Investigating Non-Periodic Solids Using First Principles Calculations and Machine Learning Algorithms

Investigating non-periodic solids using first principles calculations and machine learning algorithms

A dissertation presented
by

Ekin Doğuş Çubuk

to John A. Paulson School of Engineering and Applied Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Applied Physics

Harvard University
Cambridge, Massachusetts
April 2016

Thesis advisor

Efthimios Kaxiras

Author

Ekin Doğuş Çubuk

Investigating non-periodic solids using first principles calculations and machine learning algorithms

Abstract

Computational methods are expected to play an increasingly important role in materials design. In order to live up to these expectations, simulations need to have predictive power. To achieve this, there are two hurdles, both relating to the complexity of physical interactions. The first is the quantum mechanical interactions of ions and electrons at short timescales, which have proven difficult to simulate using classical computation. While it is now possible to model some properties and materials using first principles methods (e.g. density functional theory), accuracy, consistency and computational efficiency need to be improved to meet the demands of high-throughput materials design. The second hurdle is the difficulty of predicting the outcomes of interactions between ions at longer timescales. These interactions are central to some of the biggest mysteries in condensed matter physics, such as the glass transition.

Meanwhile, the field of machine learning and artificial intelligence has seen rapid progress in the last decade. Due to improvements in hardware, software, and methodology, machine learning algorithms are now able to learn complex tasks by mastering fundamental concepts from data. Thus, this thesis explores the applicability of machine learning to the main challenges facing computational materials design. First, as a case study, we investigate the lithiation of amorphous silicon. We show that large

unit cells need to be simulated to model lithium-silicon alloys accurately. By analyzing the geometric structures of local neighborhoods of silicon atoms, it is possible to explain the macroscopic behavior from microscopic signatures. In response to the first hurdle as discussed above, we train neural networks to reproduce energies of silicon structures and silicon-lithium alloys, which allows us to study much larger unit cells. We then explore silicon neural networks in detail, in order to explain how this specific machine learning architecture can model quantum mechanical interactions.

The following two chapters focus on the second hurdle which arises from complex ionic configurations. By studying Lennard-Jones supercooled liquids, we try to resolve two mysteries related to supercooled liquids: 1) why the dynamics are spatially heterogeneous, and 2) why the relaxation time increases super-exponentially as the temperature is lowered. Through machine learning, we can resolve the first mystery quantitatively. Furthermore, we show that the second can also be resolved in our framework, by using empirical measurements of the machine learned representation, which we call "softness". Finally, we discuss the physical meaning of softness, by comparing it to other measures and applying unsupervised learning and reduced curve-fitting models.

Contents

Listing of figures

Dedicated to my sister, Ece Doğu Çubuk

Acknowledgments

The work represented in this thesis would not have been possible without the help of many people. I have learned a tremendous amount from the people I interacted with through graduate school and I cannot list all their names here.

First of all, I would like to thank my advisor, Efthimios Kaxiras. He taught me a lot about science, presentation and writing. I admire his rigorous, creative, and honest approach to science; and I hope to be as deep and diverse as him in my academic career. I am also grateful to him for giving me just the right amount of freedom, allowing me to try things that nobody had tried before, but also stopping me when I pursued directions that did not make sense. He supported me whenever I needed personal or academic help, and I appreciate how kind and calm he has been not just to me but also to the rest of the group. I would also like to thank him for the amazing group that he built over the years. It would be hard to find so many talented and nice people working so harmoniously together, but it was possible in Kaxiras group.

Wei Li Wang introduced me to the Siesta code package and Li-ion batteries, and patiently answered all my questions as I was trying to learn about computational solid state physics. It was a pleasure to work with Brad Malone and start research into machine learning applications in our group. Brad has been a great friend, and working with him made it easier to deal with the ins and outs of academia. Wei Chen has also been a great friend, and it was a pleasure spending time in so many different

countries together. I would like to thank him for answering all my questions patiently, on topics ranging from VASP, 2D materials and to Chinese cuisine. In addition, it has been rewarding to learn from Matthew Montemore about the complex science of surface catalysis. Tian Lan and I worked on several machine learning projects in solid state physics. I was supposed to mentor him in the group, but I ended up learning a lot from him on many different topics. It was great working with him in the last two years, and I am sure he will make important contributions to whichever research direction he ends up taking. Berk Onat and I worked together on neural network potentials, and it was great working with such a hardworking and kind person. I would also like to acknowledge Alireza Ostadhossein, for all the discussions on Li-ion batteries. In addition, I would like to thank Oscar Granas, Grigory Kolesov, and Kejie Zhao, Shiang Fan, and Dennis Huang for the conversations and discussions.

I would also like to thank my committee: Ryan Adams, Alan Aspuru-Guzik, and Joost Vlassak. They were always available to meet and give me advice on science, despite their busy schedules, for which I am grateful.

Samuel Schoenholz has been my best friend for the past 10 years and our work together in the last few was the most fun part of graduate school research. I would like to thank him for introducing me to glassy liquids and soft matter physics in general. "Softness" work started with a late night conversation he and I had in Philadelphia, and we have been fascinated by it since then. Through Sam, I met Andrea Liu, his advisor. Andrea has been extremely generous with her time and attention, and I would like to thank her for all that she has taught me, and caring this much about the softness method and science in general. It has been a privilege collaborating with her in the second half of graduate school, and I hope that we can work together for many more years to come.

Amos Waterland has been a great friend since I met him at the prospective graduate students visit at Harvard University. Meeting Amos and hearing about his ambitious project was the best part about the visit, and I feel very lucky to be a part of his project. He taught me so much over the years, from computer science to applied math to physics; and he was the first person to introduce me to the outstanding potential machine learning has in contributing to physics. His dedication and curiosity for science is very admirable, and I hope to keep learning from him for the rest of my career.

I spent lots of time in University of Science and Technology in Hefei, China. I would like to thank Zhenyu Zhang and Wenguang Zhu for their hospitality. They have introduced me to different directions in computational solid state physics and materials science. I would also like to thank the many friends I have made in Hefei over the years, for teaching me about their research, Chinese culture and cuisine.

Outside of research, I was very privileged to have great friends in Cambridge that made graduate school fun, and I am grateful for their friendship. Alice Xiang kept me sane and happy, and this work would not have been possible without her support. Finally, I would like to acknowledge my father Kazım Çubuk, my mother Nezaket Çubuk, and my sister Ece Doğu Çubuk, for they made me who I am today. My parents were the ones who taught me the value of scientific thinking and curiosity from an early age, and I am very grateful for that.

1

Introduction

Computer simulations are expected to play an important role in the analysis and design of materials. Up to now, useful materials have mostly been designed in the lab, using a trial-and-error approach[2,3]. For this reason, new materials often take several decades to go from the initial design in the lab to commercial applications[4]. Materials Genome Initiative (MGI) strives to deploy new materials twice as fast, at a fraction of the cost. Computation is given an important role in the strategic plan for achieving this ambitious goal[5]. Depending on the size of the system (number of atoms) and the degree of accuracy desired, quantum mechanical (QM) calculations can take from minutes to several days to compute certain properties of certain materials. Since this timescale is significantly faster than the time it takes to synthesize, process and char-

acterize materials in the lab, many scientists are hopeful that computational methods can significantly speed up the design and deployment of new materials[6,7]. The idea is to use simulation tools in a high-throughput framework, and screen thousands of materials *in silico* quickly, before choosing a few to fabricate and test in the lab. We will call this ambition the computational revolution of materials design (CRMD).

There are two main obstacles to overcome before reaching the CRMD. First is the difficulty of doing QM calculations on a classical computer. Properties of solids are governed by the QM interactions between ions and electrons, which is computationally challenging to simulate, as discussed below. For this reason, efficient and accurate approximations to the Hamiltonian are needed. Furthermore, to apply high-throughput screening (HTS) to candidate materials, the accuracy of the calculations needs to be consistent across all candidate materials. This problem is being attacked from many different directions, including *ab initio* methods such as density functional theory (DFT)[8–10] and exchange-correlation functionals[11–16], Hartree-Fock method[17–19], GW[20–24], quantum Monte Carlo[25–28], tight binding[29–31], classical force fields[32–38], reactive force fields[39,40], and evolutionary algorithms[41,42], etc; although it is not possible for us to list all the different approaches to modeling the energy landscape ($3 \times N_{at}$ dimensional space where N_{at} is the number of atoms) of solids. Some of these methods are fast enough for certain applications and materials, and similarly some of these methods are accurate enough for calculating certain properties of some class of materials.

The second obstacle is regarding the complexity within these energy landscapes. Even if we could calculate the Hamiltonian for a solid, the metastable states and the dynamics are too complex for us to make predictive statements about which phases could be fabricated in the lab, and how the material properties would change as goes

through different metastable phases. To demonstrate the complexity of the dynamics on "simple" energy landscapes, consider the Lennard-Jones potential, or the hard/soft sphere models. Despite the simplicity of these atomic interactions, and the extensive research on this topic[1,43–47], many fundamental aspects of its dynamics are still not known from first principles. Questions like: is there a phase transition underlying the glass transition, or why do the dynamics slow down faster than exponentially in temperature as the glassy liquid is cooled down, cannot be answered from first principles (starting from the atomic interactions) even for the simplest energy landscapes. Similarly, finding stable disordered states for these simple energy landscapes is an active area of research[48,49]. For more complex atomic interactions (like DFT), finding the global minimum of a given energy landscape is a challenge, where two of the leading methods are the random guessing method the *ab initio* random structure search (AIRSS)[50], genetic algorithms[51], and basin-hopping[52].

It is clear that the two obstacles are related to two very different timescales. The first obstacle, in which we assume that the ions are fixed and the electron-ion interactions and electron-electron interactions are the cause of complexity, the timescale is between 10^{-18} to 10^{-15} seconds. For many materials properties, it is sufficient to consider only the ground state of electrons given the ions configuration. In that case, the dynamics of electrons are ignored, and only the lowest energy electron distribution is used. Regarding the second obstacle, the time scale can be as large as hours or days. For this reason, molecular dynamics is not always an option. However, even for systems where events occur in shorter timescales (like plasticity in supercooled liquids, see Chapter 4 and 5), the complexity of ion dynamics can be too complex for simple approaches free volume or local energy. In the next section, we will describe the first obstacle in more detail.

1.1 Quantum mechanical description of solids

Since quantum theory was invented, physicists tried to find approximations to the solution of complex quantum mechanical calculations. Dirac famously remarked[53]:

> The general theory of quantum mechanics is now almost complete, ….
> The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.

Today, many scientists still work on achieving the goal set by Dirac: finding appropriate approximations to the quantum mechanical interactions between atoms. Density functional theory, which tries to calculate the properties of atomic systems based only on the electron density, is the most commonly used approach today. The idea was originally proposed by Thomas[54] and Fermi[55,56], and further developed by Dirac[57]. The modern formulation of DFT can be traced back to Hohenberg and Kohn[8], who showed, among other things, that there is a one-to-one correspondence between the Hamiltonian of a system and its ground state charge density[58]. The foundation laid out by Hohenberg, Kohn, and Sham[8,9]; combined with the algorithmic, software and hardware developments in the last four decades allow us to use popular DFT packages to simulate the behavior of some solids with some accuracy.

A solid can be represented as a collection of ions and electrons, whose Hamiltonian

is given by[59]:

$$\begin{aligned}\mathcal{H} = &-\sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &-\sum_{i,I}\frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} \\ &-\sum_I \frac{\hbar^2}{2M_I}\nabla^2_{\mathbf{R}_I} + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}\end{aligned} \tag{1.1}$$

where M_I, Z_I, and $\mathbf{R}_I$ are the mass, valence charge, and the position of the ion I, respectively. m_e is the mass of an electron, $\mathbf{r}_i$ is the position of the electron i, and e is the charge of an electron. The first term is the sum of the kinetic energy of electrons. The second term is the electrostatic potential energy between the electrons, and the third term is the potential energy between ions and electrons. The last two terms are the kinetic energy of the ions and the potential energy between the ions. The many-body Schrödinger's equation is given by

$$\mathcal{H}\Psi\left(\mathbf{R}_I, \mathbf{r}_i\right) = E\Psi\left(\mathbf{R}_I, \mathbf{r}_i\right),$$

where $\Psi\left(\mathbf{R}_I, \mathbf{r}_i\right)$ is the wavefunction of the solid, and E is the energy eigenvalue. In principal, the solution of this eigenvalue equation gives us all the ionic and electronic properties of the solid. However, the number of ions and electrons ($\sim 10^{23}$) in a typical solid makes this calculation intractable. The highest ranked supercomputer on the 2015 Top500 list is TIANHE-2, which has a total memory of 1.34×10^{15} bytes[60]. The prospects are not brighter for quantum computers, as the largest quantum computers to date seem to have 7[61] to 28[62] qubits.

A commonly used approximation to get around this problem is to assume that the solid is constructed by a much smaller unit cell repeated spatially in 3 dimensions.

This can be a very good approximation for a perfect crystal, since most crystals can be constructed by tiling a 1-10 atom unit cell. This approximation however cannot directly be applied to disordered solids and crystals with defects. The unit cells need to be large enough such that the defects do not interact with their images, if periodic cells are to be used to study defects. For disordered solids, the unit cell needs to be large enough so the the properties of the solid are not significantly affected by the periodicity.

1.1.1 Born-Oppenheimer approximation

By using repeating unit cells, the problem size can be reduced from 10^{23} to tens or hundreds of atoms. Eq. 1.1 is still too difficult to solve using the computers of today. For this reason, another approximation is commonly employed, for studying the ground state of $\Psi(\mathbf{R}_I, \mathbf{r}_i)$, called the Born-Oppenheimer approximation.

m_e is much smaller than a typical ion mass. For example, the mass of a silicon atom is more than 50,000 times larger than the mass of an electron. For this reason, electrons can be expected to move much faster than ions; the dynamics of ions and the dynamics of electrons occur at different timescales. Using this observation, we assume that the many-body wavefunction can be separated into a product of the wavefunction for electrons and the wavefunction of the ions. Furthermore, we also assume that the ions can be studied as classical particles, due to their larger size. This approximation is not valid for ions whose motion needs to be studied quantum mechanically, which can be the case for light atoms like hydrogen[59].

In this picture, ions are modeled as classical particles. For any given configuration of ions, the electrons instantaneously relax to their ground state, which determines the energy of the system and the forces on the ions. Ions move according to those

forces, after which electrons find their new ground state. Thus, there is a iterative relationship between the ions and the electrons, which is directly amenable to iterative computation. We rewrite Eq. 1.1 as

$$\begin{aligned}\mathcal{H} = &-\sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,I}\frac{Z_I e^2}{|\mathbf{R}_I-\mathbf{r}_i|} \\ &-\sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_i \mathrm{V}_{ion}(\mathbf{r}_i) \end{aligned} \tag{1.2}$$

where we removed the kinetic energy of ions and the electrostatic interactions between ions, which will be considered classically as described above. In Eq. 1.2, we rewrote the interaction between ions and electrons as a potential energy only as a function of electron positions, since it is assumed that this equation will be solved for fixed ions. This is the form of the Hamiltonian that we will solve in this thesis, where only the ground state properties of the electrons are considered.

It turns out that Eq. 1.2 is still computationally too daunting within today's computational resources for more than several dozens of electrons. The reason seems to be the dimensionality of a many-body wavefunction. $\Psi(\mathbf{r}_i) = \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$ is a $3 \times N$ dimensional function which is anti-symmetric with respect to the exchange of electrons, where N is the number of electrons. A significant simplification comes from reducing the dimensionality of this problem to only 3 dimensions, as explained in the next section.

1.1.2 Hohenberg-Kohn theorems

Modern DFT has its theoretical basis on the two papers by Hohenberg, Kohn, and Sham[8,9]. Here we will discuss the first Hohenberg-Kohn theorem, which established

that there is a one-to-one correspondence between the wavefunction $\Psi(\mathbf{r}_i)$ and the charge density $n(\mathbf{r})$[8], where

$$n(\mathbf{r}) = N \int \Psi^*(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)\, \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)\, \mathrm{d}\mathbf{r}_2 \mathrm{d}\mathbf{r}_3 \ldots \mathrm{d}\mathbf{r}_N \tag{1.3}$$

We will prove this theorem by first showing that there is a one-to-one correspondence between external potential $\mathrm{V}(\mathbf{r})$ and $n(\mathbf{r})$. Suppose that two different external potentials, $\mathrm{V}_1(\mathbf{r})$ and $\mathrm{V}_2(\mathbf{r})$ leads to the same ground state charge density $n(\mathbf{r})$. We define $\mathcal{H}_1$ and $\mathcal{H}_2$ as the Hamiltonians and Ψ_1 and Ψ_2 as the ground state wavefunctions given by the potential energies $\mathrm{V}_1(\mathbf{r})$ and $\mathrm{V}_2(\mathbf{r})$, respectively. Define E_1 and E_2 as:

$$E_1 = \langle \Psi_1 | \mathcal{H}_1 | \Psi_1 \rangle$$
$$E_2 = \langle \Psi_2 | \mathcal{H}_2 | \Psi_2 \rangle$$

Invoking the variational principle:

$$\begin{aligned} E_1 < \langle \Psi_2 | \mathcal{H}_1 | \Psi_2 \rangle &= \langle \Psi_2 | \mathcal{H}_2 - \mathrm{V}_2 + \mathrm{V}_1 | \Psi_2 \rangle \\ &= \langle \Psi_2 | \mathcal{H}_2 | \Psi_2 \rangle + \langle \Psi_2 | \mathrm{V}_1 - \mathrm{V}_2 | \Psi_2 \rangle \\ &= E_2 + \langle \Psi_2 | \mathrm{V}_1 - \mathrm{V}_2 | \Psi_2 \rangle \end{aligned} \tag{1.4}$$

Similarly:

$$\begin{aligned} E_2 < \langle \Psi_1 | \mathcal{H}_2 | \Psi_1 \rangle &= \langle \Psi_1 | \mathcal{H}_1 - \mathrm{V}_1 + \mathrm{V}_2 | \Psi_1 \rangle \\ &= \langle \Psi_1 | \mathcal{H}_1 | \Psi_1 \rangle + \langle \Psi_1 | \mathrm{V}_2 - \mathrm{V}_1 | \Psi_1 \rangle \\ &= E_1 + \langle \Psi_1 | \mathrm{V}_2 - \mathrm{V}_1 | \Psi_1 \rangle \end{aligned} \tag{1.5}$$

We add Eq. 1.4 with Eq. 1.5, to get:

$$E_1 + E_2 < E_1 + E_2 + \langle \Psi_1 | \mathrm{V}_2 - \mathrm{V}_1 | \Psi_1 \rangle + \langle \Psi_2 | \mathrm{V}_1 - \mathrm{V}_2 | \Psi_2 \rangle \tag{1.6}$$

Since we assumed that both Hamiltonians has the same ground state charge density $n(\mathbf{r})$:

$$\langle \Psi_1 | \mathrm{V}_2 - \mathrm{V}_1 | \Psi_1 \rangle = \langle \Psi_2 | \mathrm{V}_2 - \mathrm{V}_1 | \Psi_2 \rangle = \int n(\mathbf{r}) \left[\mathrm{V}_2(\mathbf{r}) - \mathrm{V}_1(\mathbf{r}) \right] \mathrm{d}\mathbf{r} \tag{1.7}$$

Thus Eq. 1.6 becomes $E_1 + E_2 < E_1 + E_2$, which proves by contradiction that our initial assumption cannot be correct. Since there is also a one-to-one correspondence between external potential and ground state wavefunction, this proves that the wavefunction must be a unique functional of the charge density.

This significantly simplifies our problem, which is to calculate the properties of solids using quantum mechanics. It also shows that the quantum mechanical universe is not necessarily exponentially larger, as is often claimed, since all properties can be calculated as a functional of $n(\mathbf{r})$, which lives in $\mathbb{R}^3$. However, we will see that we do not know how to calculate properties of electrons from $n(\mathbf{r})$ directly. We can rewrite

Eq. 1.2 as:

$$\mathcal{H} = F\left[n(\mathbf{r})\right] + \int \mathrm{V}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \tag{1.8}$$

where $F\left[n(\mathbf{r})\right]$ is the sum of kinetic energy and electron-electron potential energy, and the last term is the ion-electron potential energy. $F\left[n(\mathbf{r})\right]$ is called the universal functional of charge density, since the kinetic energy of electrons and the electron-electron potential energy only depend on $n(\mathbf{r})$. The last term in Eq. 1.8 is easy to calculate since $\mathrm{V}(\mathbf{r})$ is given by the Coulomb repulsion from the ions. $F\left[n(\mathbf{r})\right]$, however, proved very difficult to find for more than 8 decades[54,56]. Thus, although all the information is contained in $n\left(\mathbf{r}\right)$, we do not know how to evaluate the hamiltonian on that charge density.

1.1.3 Single particle picture

An important step to approximate the Hamiltonian in Eq. 1.8 is to reduce the many-body wavefunction $\Psi\left(\mathbf{r}_i\right)$ to a set of single particle equations. Note that Eq. 1.8 is exact. By describing $\Psi\left(\mathbf{r}_i\right)$ as a set of single particle equations, however, the calculation becomes an approximation. The charge density then becomes:

$$n(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \tag{1.9}$$

where $\phi_i(\mathbf{r})$ are the single particle equations (orbitals). Then $F\left[n(\mathbf{r})\right]$ can be written as

$$F\left[n(\mathbf{r})\right] = T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})], \tag{1.10}$$

where $T_s[n(\mathbf{r})]$ is the kinetic energy of the non-interacting electrons, $E_H[n(\mathbf{r})]$ is the classical electrostatic potential energy, and $E_{xc}[n(\mathbf{r})]$ is the exchange-correlation en-

ergy. $E_{xc}[n(\mathbf{r})]$ is defined to make $F[n(\mathbf{r})]$ approximately accurate. For example, $T_s[n(\mathbf{r})]$ does not include all the kinetic energy since these electrons do interact via the charge density, and all the kinetic energy that was not included in $T_s[n(\mathbf{r})]$ is included in $E_{xc}[n(\mathbf{r})]$. Then we write the single particle Schrödinger's equation as:

$$\left[-\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}}^2 + \mathrm{V}^{eff}(\mathbf{r}, n(\mathbf{r}))\right]|\phi_i\rangle = \epsilon_i\phi_i(\mathbf{r}), \tag{1.11}$$

where on the left hand side, the first term is the kinetic energy of an electron and the second term is the effective potential energy the electron feels due to the ions, other electrons, and the exchange correlation function:

$$\mathrm{V}^{eff}(\mathbf{r}, n(\mathbf{r})) = \mathrm{V}(\mathbf{r}) + e^2\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' + \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})}, \tag{1.12}$$

where the last term is the variational functional derivative of the exchange-correlation functional. These single particle states, $\phi_i(\mathbf{r})$ are called Kohn-Sham (KS) orbitals. They are not electrons, since they were derived by first assuming no interactions. However, in many materials, they do maintain several of the key properties of the actual electrons[59]. Eq. 1.12 gives the single particle energy of an "electron", which includes its kinetic energy and potential energy due to ions (attractive) and the charge density (repulsive). The charge density, in turn, is made of all the electrons. Thus, this approximation naturally lends itself to an iterative solution: 1) start with initial guesses for the KS orbitals, 2) calculate $n(\mathbf{r})$ and V^{eff} from KS orbitals, 3) solve the eigenvalue equation 1.12, whose eigenvectors will be the new KS orbitals, 4) go back to item (2). This iteration is continued until the KS orbitals calculated from item (3) are the same as the KS orbitals that were used in item (2), at which point the charge

density is converged.

We have omitted the discussion of how the exchange-correlation functional is constructed. There is no exact form for the $E_{xc}[n(\mathbf{r})]$ yet, and this term contains all the approximations of the modern DFT calculations. Instead, empirical functionals have been used that give reasonably accurate results for certain materials.

1.2 Opportunities for machine learning methods

As described in the previous section, although DFT is considered an *ab initio* method, it has an empirical term for exchange-correlation functional. Given the complexity of approximating this functional, combined with the massive DFT datasets that are being available in recent years[2,63,64]; machine learning can be a viable approach to improving the speed and accuracy of DFT. There are many different approaches to machine learning DFT energies. Most common approach to date is to learn the mapping from $\mathbb{R}^{3N_{at}}$ to potential energy. This approach is often limited to specific solids, with at most 2-3 types of atoms. Examples include Behler and Parrinello's neural network architecture[65], Gaussian Approximation Potentials[66], and linear, sparse models by Seko et al.[67]. Another approach is to use machine learning for the universal functional in DFT[68], $F\left[n(\mathbf{r})\right]$. Since this functional is the same for any charge density, it can in principle be applied to any DFT calculation, which makes it less restrictive than the previous approach.

On the other hand, machine learning has not been employed in literature to address the complexity that arises from ionic configurations. Due to the complexity of ionic configurations, the glass transition and glassy dynamics are often studied in high dimensions, as a mean field approximation, where the structure is not important[46,47]. However, in three dimensions, structure is much more influential on the

dynamics, and the complexities of the structure need to be incorporated into the theoretical models. For this reason, we have investigated applications of machine learning in glassy liquids in Chapters 4. We show that a predictive model of rearrangements in disordered liquids is inherently very complex, previously used physical measures are not sufficient. However, by using machine learned representations, it is possible to incorporate the structural complexities in theoretical models.

There are many physical processes that have been difficult to study due to the complex arrangement of atoms. Examples include non-equilibrium aging of glasses, modeling of glass forming ability, materials failure like fracture, necking, and shear bands, etc. We believe that a data-scientific approach that includes the structural aspects of disordered solids can provide new insights.

1.3 Overview

In Chapter 2, we study the lithiation of amorphous silicon, using DFT and geometric analysis. We use this system as an example for the two obstacles described above: 1)DFT calculations have to be done on large systems with as many as 350 atoms, and 2) the disordered nature of atomic configurations need to be studied to understand macroscopic behavior observed in experiments[69]. In Chapter 3, we present a neural network based empirical potential for silicon atoms. We investigate the representations of silicon atoms in the neural network, to understand how machine learning can be used to reproduce DFT calculations.

In Chapter 4, we focus on a much simpler energy landscape: the Lennard-Jones supercooled liquid. We show that, although the atomic interactions are very simple, building predictive models are difficult, and machine learning methods are required to make correct predictions[70]. In Chapter 5, we use the representation from

Chapter 4 to study the glassy dynamics of Lennard-Jones liquid. We show that by combining machine learned representations with theoretical physics tools, one can explain the mysterious spatially heterogeneous dynamics and super-exponential relaxation times[71]. Finally, in Chapter 6, we compare our machine learned representation to other measures used for similar applications. We show that the machine learned model is predictive, whereas other measures are not. By using unsupervised learning and reduced curve-fitting models, we explore the physical meaning of the machine learned representation[72].

2

Theory of structural transformation in lithiated amorphous silicon

2.1 Overview

Determining structural transformations in amorphous solids is challenging due to the paucity of structural signatures. The effect of the transitions on the properties of the solid can be significant and important for applications. Moreover, such transitions may not be discernible in the behavior of the total-energy or the volume of the solid as a function of the variables that identify its phases. These issues arise in the context of lithiation of amorphous silicon (a-Si), a promising anode material for high-energy-

density batteries based on lithium ions. Recent experiments suggest the surprising result that the lithiation of a-Si is a two-phase process. Here, we present first-principles calculations of the structure of a-Si at different lithiation levels. Through a detailed analysis of the short and medium-range properties of the amorphous network, using Voronoi-Delaunay methods and ring statistics, we show that a-Li$_x$Si has a fundamentally different structure below and above a lithiation level corresponding to $x \sim 2$.

2.2 Introduction

Materials are slated to play a key role in addressing future challenges in energy use. As a case in point, high-energy density batteries, often based on Li-ion technology, are important for many mobile applications. Increasing the capacity and durability of Li-ion batteries is necessary for enabling solutions in various energy sectors, such as transportation[73,74]. Silicon is a promising anode material for Li-ion batteries with a high theoretical capacity, because it can absorb more than 4 Li atoms per Si atom, as compared to the capacity of graphite, a common anode material, which absorbs up to 1 Li atom per 6 C atoms[75,76]. This has fueled an intense research effort for integrating Si in Li-ion battery design[75–79]. Understanding the lithiation/delithiation process of Si anodes is very important for designing the next-generation Li-ion batteries, because at the high lithiation limit the material expands by up to 400% and the anode fractures and cannot sustain many charging cycles. The fracture behavior depends sensitively on the lithiation/delithiation process[80].

The lithiation reaction of crystalline Si (c-Si) has been studied extensively and has been described in detail, both theoretically and experimentally[81–90]. c-Si becomes amorphous after the first cycle of lithiation and delithiation. Much less is known about these processes in amorphous Si (a-Si), which has better lithiation character-

istics than c-Si[80]. Lithiation of c-Si is a two-phase reaction, in which c-Si is lithiated layer by layer[81,82,87,91]. Lithiated Si is separated from pristine c-Si by a sharp ($\sim$ 1 nm thickness) reaction front, which moves into c-Si as the reaction progresses. Recent experiments suggest that the lithiation of a-Si is also a two-phase mechanism, contrary to previous understanding[80,92]. These experiments indicate a clear boundary between lithiated and pristine a-Si as the reaction progresses. This is quite surprising given the lack of crystalline order in a-Si, although the thickness of the reaction front may not be as sharp as in c-Si. In addition, one experimental study has proposed that as the reaction front progresses into pure a-Si, the lithiated Si has relatively constant Li concentration of $x \approx 2.5$[92]. After most of the a-Si is lithiated to the concentration of $x \approx 2.5$, a second step of the reaction occurs, where the concentration of the a-Li_xSi increases from $x \approx 2.5$ to $x = 3.75$. The experimental results are not conclusive on whether this second step is a single-phase or two-phase reaction[92].

In order to resolve these issues, and to provide a comprehensive picture of the a-Si lithiation processes, it is necessary to describe the relevant atomistic features with an unbiased and reliable theoretical model. It is also important to include in the model enough degrees of freedom to capture the complexity of the amorphous network and the changes it undergoes during the lithiation/delithiation process. Previously reported atomistic models of a-Li_xSi with first-principles calculations used small unit cells with periodic boundary conditions to describe the structures at different Li content (x values)[93–95]. The relatively small size of the unit cells imposes severe limitations on how well the model captures the nature of the amorphous solid. Large unit cells of a-Li_xSi have recently been studied using empirical potentials of the type second nearest neighbor modified embedded atom method (2NN MEAM)[96] and reactive force field (ReaxFF)[97–99].

2.3 Results

We present detailed structural and geometric analysis of a-Li$_x$Si for a very wide range of concentrations, $x = 0 - 4.25$. To improve the statistics, for each x value, we have investigated 10 different structures. Each one of our structures has 64 Si atoms, and a corresponding number of Li atoms for different x values. The largest unit cell, corresponding to $x = 4.25$, contains 336 atoms. To achieve a detailed understanding of the effect of varying x on medium-range order, we analyzed ring statistics and the geometric properties of the Delaunay triangulation. Our results suggest that a-Li$_x$Si goes through a structural transformation during lithiation until about $x \approx 2$, after which the short-range and medium-range order stays relatively constant in several important aspects. We discuss how this analysis can help explain the lithiation reaction and crystallization of a-Si observed in experiments. We also offer some predictions on why limiting the range of lithiation in batteries with a-Si anodes could significantly reduce the structural degradation of the electrodes.

To obtain reliable structural models, we used first-principles calculations within density functional theory with the SIESTA code[100] and the PBE exchange-correlation functional[101]. The Kohn-Sham orbitals are represented by a double-ζ plus polarization basis set, with an energy cutoff of 70 Ry. Structural relaxations were considered converged when the magnitude of the force on each atom was smaller than 0.04 eV/Å. Our starting point was an amorphous bulk structure consisting of 64 Si atoms with no coordination defects to model the a-Si electrode before lithiation[102,103]. To model the lithiation of a-Si, we inserted Li atoms in selected positions of the a-Si structure, and relaxed the structure in terms of both atomic positions and lattice vectors. At each step, Li atoms were placed in tetrahedral-like ("Td-like") positions in a-Si. To find

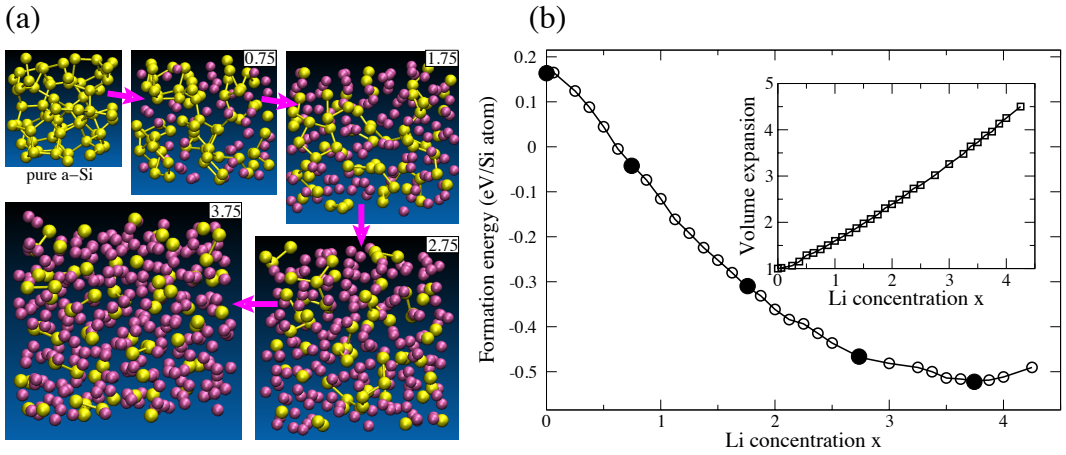


Figure 2.1: (a) Representative unit cells of the structures analyzed (yellow spheres are Si atoms and purple spheres Li atoms). (b) Average formation energy as a function of Li concentration. Filled circles correspond to the concentrations at which the structures are shown in (a). The inset in (b) shows the volume relative to the initial volume of a-Si. The results are average values over 10 structures at each value of x. The standard deviation for each point is smaller than the marker size.

these positions, we identify Si atoms with the highest number of Si neighbors at each Li concentration. We then pick one of these Si atoms and one of its Si neighbors randomly and insert the Li atom along the extension of the chosen Si-Si bond at a distance approximately equal to the bond, and relax the structure as described. At each x value, we constructed 10 different structures based on structures at the previous x value, with the position of each Li atom chosen randomly among the available positions defined above. There are many such positions in the pure a-Si network, which gives a broad range of initial structures for the first lithiation step. We show five of the structures from our calculations in Fig. 2.1(a). The amount of Li was increased in steps of $\Delta x = 0.125$. We arrived at the value of Δx after experimenting to determine that this choice does not affect the resulting structure. For small values of x,

our method of lithiation is equivalent to the protocol used by Chevrier *et al.*[93].

The formation energy $E_f(x)$ of a-Li$_x$Si per Si atom was calculated by the expression

$$E_f(x) = (E_{\mathrm{Li}_x\mathrm{Si}} - E_{\mathrm{Li}}N_{\mathrm{Li}} - E_{\mathrm{Si}}N_{\mathrm{Si}})/N_{\mathrm{Si}} \tag{2.1}$$

where $E_{\mathrm{Li}_x\mathrm{Si}}$ is the energy of the lithiated a-Si system, E_{Li} is the energy of one Li atom in BCC bulk Li, E_{Si} is the energy of one Si atom in c-Si, and N_{Li} and N_{Si} are the number of Li and Si atoms in the system, respectively. $N_{\mathrm{Si}} = 64$ for all of the systems considered here. In Fig. 2.1(b) we show $E_f(x)$ of a-Li$_x$Si per Si atom as a function of Li concentration x. There is a shallow minimum at $x = 3.75$. There is no evidence for any structural phase change as a function of x in the formation energy and the volume of the model.

Since lithiation of c-Si and delithiation of c-Li$_{15}$Si$_4$ are two-phase reactions, values of the plateaus in the voltage V observed in the experiments can be calculated from the formation energies, $E_f(x_1), E_f(x_2)$ of the initial and final phases and the Li concentrations, x_1, x_2 in these phases[104] :

$$V = -\frac{1}{e}\,\frac{E_f(x_2) - E_f(x_1)}{x_2 - x_1} \tag{2.2}$$

where e is the charge of the electron. The results and a comparison to experiment and earlier calculations[104] are shown in Table 1. Our results are in very good agreement with experiment. The improvement over previous theory calculations is due to the fact that we use much larger unit cells to model a-Si, which reduces the translational order imposed by periodic boundary conditions and gives a better estimate for the formation energy of the amorphous structures. This level of agreement strongly suggests that our models capture the essential nature of the amorphous pure and lithi-

	c-Si $\rightarrow$ a-$Li_{3.75}Si$			c-$Li_{15}Si_4$ $\rightarrow$ a-Li_2Si		
	$E_f(0)$	$E_f(3.75)$	V	$E_f(3.75)$	$E_f(2)$	V
Experiment[104]			0.12			0.40
Earlier theory[104]	0	-0.766	0.20	-1.106	-0.543	0.32
Present theory	0	-0.530	0.14	-1.106	-0.372	0.42

Table 2.1: Formation energies E_f (in eV) per Si atom of amorphous phases and calculated plateaus in the voltage V (in Volts) for phase transitions during lithiation and delithiation.

ated Si structures.

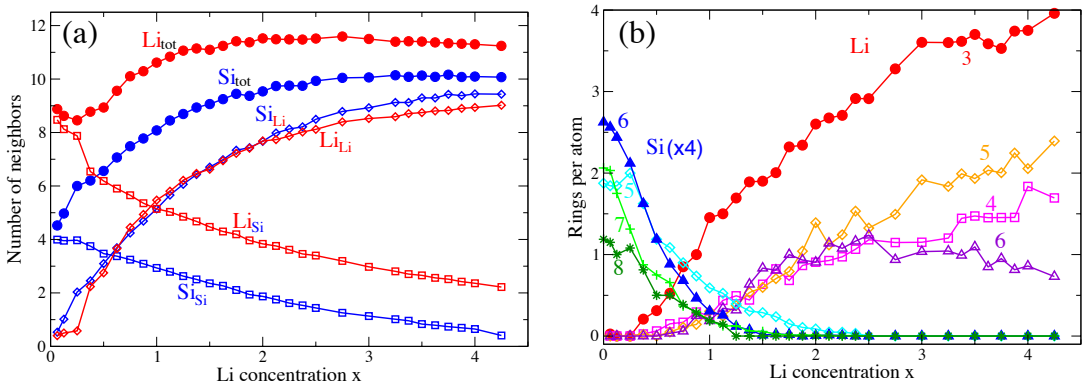


Figure 2.2: (a) X_Y denotes the number of nearest neighbors of type Y for atom type X (X, Y = Si, Li), as a function of Li concentration x. (b) The number of Si and Li rings per atom. (the number of Si rings is multiplied by 4 to put on a scale similar to the number of Li rings). In each case the number represents an average over 10 samples; the standard deviation for each point is smaller than the marker size.

We next examine the possibility of identifying a structural transformation as a function of lithiation by analyzing the short- and medium-range order of the lithiated structures. To analyze the short-range order at different lithiation stages we considered the number of Li and Si neighbors for both Li and Si atoms, shown in Fig. 2.2(a). The number of neighbors was determined from total and partial pair distribu-

tion functions that were calculated for all lithiation concentrations (Fig. S1), which are in agreement with previous results[93,94]. To calculate the number of nearest neighbors of a certain kind, we fit a gaussian to the nearest neighbor peak of the corresponding pair-distribution function and integrate it. With this procedure we can account for the fact that the average distance of neighbors and cut-off range for counting a neighbor can change as a function of lithiation. For example, the average Li-Li (Si-Si) bond distance decreased (increased) with increasing x. This was also reported by Kim *et al.*[94] and attributed to the fact that Li-Li bonds become stronger while Si-Si bonds become weaker with increasing lithiation . There are still some Si-Si bonds remaining even at $x = 4.25$, although at that level of lithiation most Si atoms are only bonded to Li atoms. It is interesting that the total number of neighbors for Si or Li changes rapidly until $x = 2$, and then saturates. This is an example of a structural property that has a different behavior before and after $x \approx 2$. This finding is similar to a previous first-principles result, where it was shown that the number of effective neighbors of Li increases with constant rate until $x = 2.2$, and the rate of increase is reduced past this value[104]. This finding was used to explain the voltage drop observed as a function of lithiation at $x = 2.2$[104]. Our results show that this is probably related to the fact that lithiation of a-Si is a two-stage reaction, and the voltage drops in the second stage. However, we emphasize that the behavior of the number of neighbors cannot be used to substantiate a two-stage reaction; additional arguments based on medium-range order are needed to make this assertion, as discussed in more detail next.

A first step toward describing medium-range order is ring statistics[105,106]. Other recent works also used ring statistics in a-Li_xSi to investigate the atomic-level mechanisms of lithiation[95]. In that work, ring statistics were reported for only 4 different

lithiation levels and there was no distinction made between rings formed by Li or Si. We present the ring statistics separately for Li and Si atoms which reveals interesting patterns regarding medium-range order, at finely sampled x values in Fig. 2.2(b). We implemented the ring counting method with the shortest-path criterion as developed by Franzblau[105]. Any two atoms were considered to be bonded if they are within each others first neighbor peak in the corresponding pair-distribution function. At $x = 0$, the a-Si structure consists of mostly 6-fold Si-rings, followed by 7-fold Si-rings and 5-fold Si-rings, in agreement with previously reported results[105]. Si rings of all sizes decrease in number with increasing lithiation, since Li atoms break more and more Si-Si bonds. For $x \geq 0.5$, 5-fold Si-rings become dominant. For $x \geq 2$, the Si rings have been eliminated in the lithiated structure. For Li atoms, 3-fold rings are always the most prevalent. At low values of x all existing Li rings are of size 3. The ratio of 3-Li-rings decrease with increasing lithiation, as larger rings form. Starting around $x \approx 2$, the ratio of 3-fold Li-rings to the total number of Li-rings saturates at $\sim 40\%$ and remains essentially constant for $2 \leq x \leq 4.25$.

The ring statistics already reveal strong hints of a structural transformation until around $x \approx 2$, with the number of Si-rings of any size completely disappearing and the number of 3-fold Li-rings dominating and becoming relatively constant (as a fraction) beyond this level of lithiation. Additional, and stronger, evidence for this is obtained from Voronoi-Delaunay methods, which are quite powerful in analyzing non-crystalline structures[107–110]. In this type of analysis, the unit cell is partitioned into tetrahedra, where positions of the atoms correspond to the vertices of these tetrahedra. To analyze the local structure of a-Si, a well-defined geometric measure for tetrahedra formed by Si atoms was proposed[110], in which four nearest neighbors of a Si atom were considered to be the vertices of a tetrahedron. By comparing how sim-

ilar each tetrahedron is to a regular tetrahedron, a local shape measure was used to differentiate between different structural phases of Si. We will call this shape measure the "diamond-measure" ($\chi_{\rm D}$): the value of $\chi_{\rm D}$ is minimized when the shape of the tetrahedron formed by the atoms approaches that of a regular tetrahedron (which corresponds to $\chi_{\rm D} = 0$, as is the case for c-Si).[110].

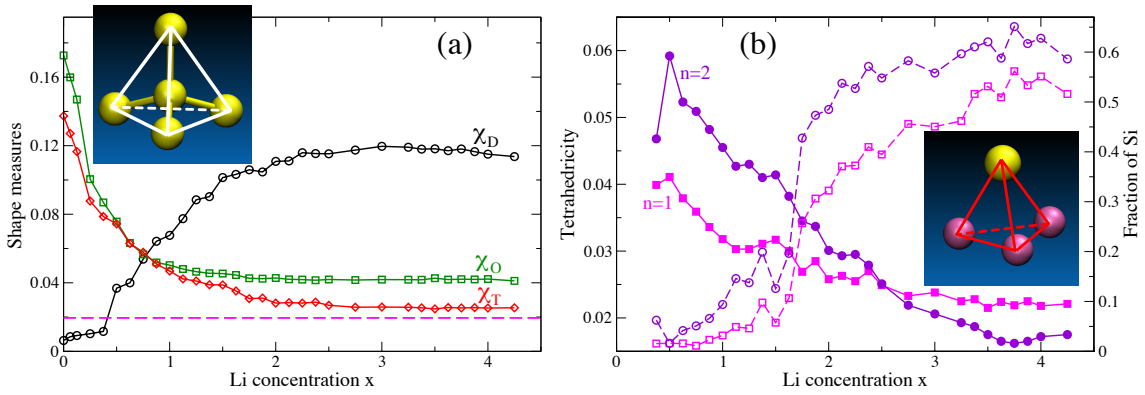


Figure 2.3: (a) Values for shape measures of tetrahedral order during lithiation: Lower values correspond to higher order (smaller deviation from ideal shape). The dashed line represents the cutoff for good tetrahedricity, as explained in text. The inset shows a regular tetrahedron formed by four nearest neighbors of a Si atom ($\chi_{\rm D} = 0$). (b) Solid markers represent the values of tetrahedricity of Delaunay tetrahedra with one or two Si atoms. Hollow markers represent the fraction of Si atoms that are in at least one good Delaunay tetrahedron with only one Si (hollow squares) or up to two Si atoms (hollow circles). The inset shows a good tetrahedron with one Si atom and three Li atoms. The numbers plotted represent median values over all tetrahedra and all samples we investigated at each value of Li concentration.

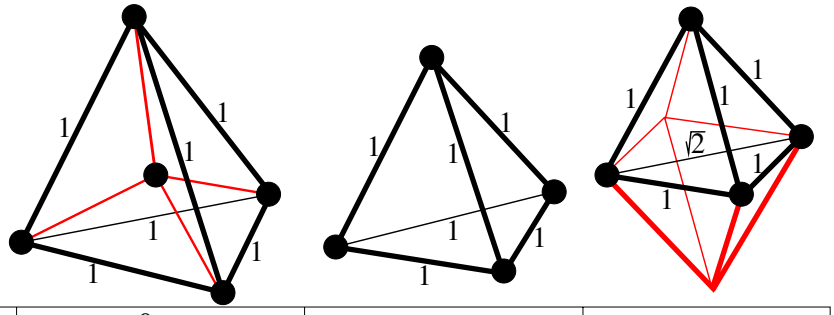
In Fig. 2.3(a), we present the median values of different shape measures as a function of lithiation. These are the median values of shape measures of all tetrahedra in the structures of corresponding x. The larger the shape measure, the less similarity the tetrahedra have to the corresponding shape. At $x = 0$, $\chi_{\rm D}$ is very small, since atoms in a-Si have local environments that are very similar to those of c-Si. The $\chi_{\rm D}$ value of a-Li$_x$Si increases with x, as expected, since the local environment of Si atoms

is severely disrupted from tetrahedral symmetry as Li atoms break the Si-Si bonds in the amorphous network. Although it is clear that the structure loses its diamond-like tetrahedral symmetry with increasing x, it is not clear from χ_{D} alone whether or not any other type of order exists. To address this issue, we used several other shape measures, by analogy to the diamond measure χ_{D}. These are referred to as "tetrahedricity" (χ_{T})[107–109] and "octahedricity" (χ_{O})[107,108]. In these cases, unlike χ_{D}, the tetrahedra are formed by Delaunay triangulation and they do not have an atom at the center. χ_{T} is a measure of how regular each Delaunay tetrahedron is. In the case of χ_{O}, the tetrahedra formed by atoms are not regular, but correspond to a quarter of a regular octahedron. The expressions for these shape measures are given by

$$\chi_{\mathrm{I}} = \frac{1}{N_p \cdot \langle e \rangle^2} \sum_{i<j} \left(\frac{e_i}{\lambda_i} - \frac{e_j}{\lambda_j} \right)^2 \tag{2.3}$$

for I = T, O or D and $e_{i,j}$ are the lengths of the 6 edges, $\langle e \rangle$ is the average edge length of the tetrahedron and N_p is the number of all possible edge pairs. For I = T or D, $\lambda_{i,j} = 1$, while for I = O, $\lambda_{i,j} = \sqrt{2}$ for i and j that refer to the longest edge of the tetrahedron. The "perfect" tetrahedra corresponding to each shape measure is shown in Fig. 2.4. The results for all these measures of medium-range order are shown in Fig. 2.3(a). The χ_{D} order vanishes with increasing x, while χ_{O} and χ_{T} order starts to emerge, with both of these quantities acquiring very low values (corresponding to very small differences of the tetrahedra from their ideal shapes). The transformation from one kind of order (low χ_{D}) to a different kind of order (low χ_{T} and χ_{O}) has been completed at around $x \approx 2$, with almost no change beyond this value: in the range $2 \leq x \leq 4.25$, the χ_{T} and χ_{O} shape measures are all roughly constant. Although we only show the median values at different x, it is important to note

that the distributions of χ_T and χ_O are also very similar for each x value between 2 and 4.25. This is another indication of a-Li$_x$Si undergoing a structural transformation with increasing x until around $x = 2$.



χ_D	0		
χ_T	0.14	0	0.05
χ_O	0.18	0.03	0

Figure 2.4: Schematic of the perfect tetrahedra for the each shape measure (from left to right): χ_D,χ_T and χ_O. These denote the shape measure for diamond-measure, tetrahedricity, and octahedricity, respectively. The table includes the values of these tetrahedra for different shape measures. Lower values indicate higher similarity to the corresponding perfect shape.

To establish what would be a good level of order, we analyzed the tetrahedricity of the crystalline $Li_{15}Si_4$ structure and found that the upper limit for good tetrahedra in the histogram of χ_T is at $\chi_T = 0.020$ (shown as a dashed purple line in Fig. 2.3(a)). This upper limit, established similarly by identifying the first minimum in the histogram of χ_T, was found to be 0.018 for a thermally excited face-centered cubic crystal[111]. Accordingly, we will use the value $\chi_T = 0.020$ to distinguish structures of high or low level of order. During lithiation, the median χ_T value approaches this limiting value. More insight can be gained by examining the tetrahedricity of special classes of tetrahedra, defined by their composition. This is shown in Fig. 2.3(b), where χ_T for tetrahedra containing only one Si atom and 3 Li atoms, or 2 Si atoms and 2 Li atoms are shown separately. This analysis reveals that at higher lithiation

(for $x \geq 2.25$), most tetrahedra have one Si atom and 3 Li atoms. Specifically at $x = 3.75$, 32% of the tetrahedra have no Si atoms, 61% have one Si, and 7% have 2 Si atoms. The χ_T value of simplexes with only one or two Si atoms is even lower than for the ones with no Si atoms. Especially tetrahedra with 2 Si atoms have median χ_T values below the limiting value of 0.020 for $x \geq 3.25$, with a global minimum at $x = 3.75$. To understand the implications of these results, especially in what concerns the phase transition from a-$Li_{3.75}Si$ to c-$Li_{15}Si_4$ which is observed experimentally[112,113], we present in Fig. 2.3(b) the fraction of Si atoms that belong to a good tetrahedron (one with $\chi_T \leq 0.020$) as a function of lithiation. In this plot, hollow circles indicate the fraction Si atoms in a good tetrahedron and the hollow squares represent the fraction of Si atoms in a good tetrahedron with no other Si atoms. In the c-$Li_{15}Si_4$ structure every Si atom belongs to a good tetrahedron with 3 Li atoms (with bond lengths 2.72 Å and 2.65 Å for Li-Si and Li-Li pairs, respectively)[114]. For this tetrahedron, $\chi_T < 10^{-3}$. Thus, the fraction of Si-containing tetrahedra included in Fig. 2.3(b) gives an estimate of the percentage of Si atoms that have a local environment that is geometrically similar to that found in c-$Li_{15}Si_4$. The fraction of hollow squares peaks at 56% in Fig. 2.3(b), which is the fraction of Si atoms that are in a good tetrahedron with 3 other Li atoms. The fraction of hollow circles peaks at 66%, since this value includes tetrahedra with one or two Si atoms. Both peaks occur at $x = 3.75$, the concentration at which the transition from a-$Li_{3.75}$ Si to c-$Li_{15}Si_4$ occurs. This shows that in a-$Li_{3.75}Si$ the majority of Si atoms are already in a local environment very close to the one found in the crystalline phase of the same composition, which would make the transition between the amorphous and crystalline structures kinetically easy. This also agrees with the experimental observation that the phase transition does not involve large scale atomic migration nor phase separa-

tion[113].

2.4 Discussion

The results discussed so far give very strong indications that there are two distinct phases involved in the lithiation of a-Si: one phase below Li content $x \approx 2$ and a different phase for higher Li content. At $x = 3.75$ the local order in the amorphous phase is quite close to the atomic environment for Si atoms in the crystalline $Li_{15}Si_4$ structure, making the transition between the two kinetically easy. The essential difference between the structures below and above $x \approx 2$ can be gleaned by a careful look at the structures of Fig. 2.1(a): up to $x \approx 2$ there is a well defined network of covalently bonded Si atoms, whereas this network has been dissolved at higher Li concentrations, and all that remains are disconnected clusters of Si atoms with few covalent bonds among them, embedded in a dense, amorphous structure of Li atoms.

Finally, it is worth exploring the implications of these findings for the operation of Li-ion batteries that use Si as the anode. From what we find, it can be concluded that if the lithiation/delithiation cycle is restricted to the range $2 \leq x \leq 4.25$, the Si-Li structure will always remain in the same structural phase. Specifically, the high-lithiation phase will consist mostly of small Si clusters, whose size decreases upon further addition of Li. From these considerations it appears that the operation of the battery within this range would be beneficial in terms of avoiding anode degradation upon charging/discharging cycles.

3

Representations in neural network based empirical potentials and force fields

3.1 Overview

The structural and mechanical properties of crystals, glasses and biological macromolecules are determined by local interactions between atoms. These interactions ultimately derive from the quantum nature of electrons, which can be prohibitively expensive to simulate. Machine learning has the potential to revolutionize materials modeling, due to its ability to find patterns in large volumes of data. So far, this type of approach has been used as an approximation tool to speed up computationally ex-

pensive calculations. We argue that machine learning models trained on physical systems can be used as more than just approximations since they had to "learn" physical concepts in order to reproduce the labels they were trained on. We use dimensionality reduction techniques to study in detail the representation of silicon atoms at different stages in a neural network, which provides insight into how a neural network learns to model atomic interactions.

3.2 Introduction

Although the ultimate basis of atomic interactions in solids is the quantum mechanical behavior of valence electrons, classical interatomic potentials are widely and successfully used to describe many properties of materials. Conventional approaches to deriving interatomic potentials are based on physically motivated functional forms. For example, the embedded atom method (EAM), commonly used to model atomic interactions in metals, is based on the intuition that the electrons responsible for the interactions are spherically distributed around the atoms, and hence the charge density of a metallic solid is largely isotropic[115]. These implied approximations determine the success *and* the limitations of the EAM. Analogous concepts based on physical intuition are used to construct interatomic potentials for covalently or ionically bonded solids and molecules. It is not easy to generalize this approach to systems of mixed character, or systems that contain a large number of different elements. As a response to this challenge, several ML-based empirical potentials have been proposed in the last decade[65,66]. Since ML algorithms use flexible functional forms, any kind of atomic interaction that is present in the data can, in principle, be "learned" with sufficient training. Through the use of clever molecular dynamics (MD) schemes[116–119] and careful construction of training sets, ML-based potentials can allow for much

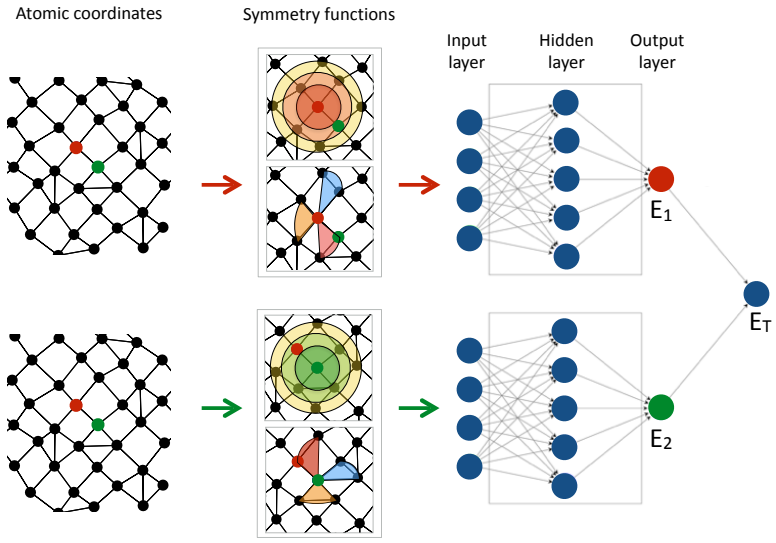


Figure 3.1: Schematic of the NN used to construct the interatomic potential, with contributions from many atoms (two atoms are highlighted red and green), starting with atomic coordinates which are transformed to symmetry functions of radial (represented as circles) and angular (represented as triangular funnels) character.

longer and larger simulations with accuracy comparable to that of first-principles quantum mechanical approaches like density functional theory (DFT). Due to the complexity and the flexibility afforded by ML algorithms, it is difficult to extract information on how they can model complicated systems. For this reason, ML-based potentials are often viewed as "black boxes" that approximate the target values on which the algorithm was trained and not as a tool that can produce conceptual insights.

Here we propose an approach to extract what a NN has "learned" in order to predict total energies as a function of the atomic positions, and to elucidate how this was accomplished. Specifically, we consider Si as a model system and explore its behavior in several different solid and liquid phases. We begin by training a NN potential to reproduce total energies as a function of atomic configuration that were originally

obtained through DFT calculations. We use the NN architecture proposed by Behler and Parrinello[65], which has been employed to successfully construct empirical potentials for semiconductors[120,121], metals[122,123], and water clusters[124] (see Methods for more details). The input to the NN are symmetry functions of the atomic positions, which describe each atom's local environment (Fig. 3.1). We then explore the internal characteristics of this potential by investigating the representation of the Si crystal as it passes through the NN architecture. This allows us to study and visualize the energy landscape of Si in an unbiased way, since we do not impose any physical intuition in the construction of the potential. By focusing on the interaction of an individual Si atom with its local environment, we can present the complicated many-body nature of the interaction in a simple and intuitive manner.

3.3 Methods

Our training set is composed of 64-atom Si cells in four different phases: cubic-diamond (*CD*), *β-tin*, *R8* and liquid (*L*). Data is collected by running MD at temperatures of 300, 600 and 900 K for the solid phases and up to 2,500 K for the *L* phase. Additional data is prepared by random distortions of the crystal structures. All the unit cells are relaxed at zero pressure. The MD simulations are performed within the Vienna *ab initio* simulation package (VASP)[125], using a projector-augmented wave (PAW) potential[126,127]. We use the generalized gradient approximation proposed by Perdew, Burke, and Ernzerhof as the exchange-correlation functional[11]. The k-point mesh is optimized for an energy convergence of 0.5 meV/atom and a stress convergence of 0.1 kbar. The plane wave energy cutoff was set to 300 eV. In order to reduce the correlation between data points, we thin the MD data by using one out of every 100 consecutive structures from the MD simulations at 300K, and one out of every 20 struc-

tures from higher temperature MD simulations. After the thinning, 23,000 64-atom structures from many temperatures and phases and their DFT energies are used to train and validate the NN. Of this data, 18,000 structures were randomly chosen as the training set and the rest were used as an independent test set. We use the symmetry function set reported in Ref.[128] as our descriptors, which consists of 8 radial and 22 angular symmetry functions for an input vector of size 30 for each atomic contribution. We reduce the number of symmetry functions to 28 since 2 of the angular functions are linear combinations of other angular symmetry functions. Since each structure has 64 atoms, our data set has a total of $(23,000 \times 64) \sim 1.5$ million local neighborhoods.

The total energy of a system is given by the sum of individual atomic contributions to the total energy, each atomic contribution being represented by a NN[65]. In the case of a system in which all atoms are the same, each NN has the same numerical weights and biases, and they are updated in a concurrent way as the network is trained. To train the NN, total energies of Si cells containing a relatively large number of atoms, obtained from MD simulations using DFT-derived forces for the dynamics, are used as labels. The non-linear transformations applied to the symmetry functions by a trained NN can reproduce their energies almost as accurately as DFT (within 3-6 meV/atom). The final output of an individual Si atom's NN is the energy contribution E_i of that Si atom's local neighborhood, which is a linear combination of the numbers that come out of the N_n nodes in the last hidden layer. Thus, these N_n numbers that the last hidden layer outputs can be thought of as the representation[129] of a Si atom's local neighborhood within the NN, after which a simple linear operation gives E_i, as illustrated in Fig. 3.1. The calculation of symmetry functions and the subsequent application of the hidden layers transform the Cartesian coordinates of

atoms so that the resulting N_n numbers in the last hidden layer is a convenient representation for predicting the energy.

The NNs are trained with the Torch 7 software package[130], using a GPU. In examining different NN architectures, we explore 1 to 3 hidden layers with the number of hidden nodes N_n ranging from 10 to 60 in each layer and pick the topology that performs best on the test set. The lowest test error is achieved using a NN with 2 hidden layers and $N_n = 60$ nodes in each layer. This network achieves a root mean squared error (RMSE) of 5.1 meV per atom on the training set, and 5.7 meV/atom on the test set. This is comparable to the RMSE of 4-5 meV/atom for the training set and 5-6 meV/atom for the test set for Si reported in the literature using the same network architecture[65]. As a simple baseline, the mean predictor, which predicts the mean value of the energy labels in the test set for every structure, achieves an RMSE of 182 meV/atom. A separate NN was trained only on the L and CD phases, in order to focus on the structural transition between them. With less variety of neighborhoods in the data set, the NN can focus on these two phases, with lower RMSE on the training and test sets. The NN with the same topology of 2 hidden layers of $N_n = 60$ achieves a RMSE of $\sim$ 2 meV per atom on the training set and $\sim$ 2.4 meV per atom on the test set.

3.4 Results

As a first attempt to visualize atomic neighborhoods we use PCA[131] as a dimensionality reduction algorithm on the symmetry function representation and the hidden layer representation. In Fig. 3.2(a), we plot the symmetry function representation in the first two principal components (PC1, PC2). Each point represents a 64-atom Si cell, color coded for the different phases. Structures with CD order in the nearest

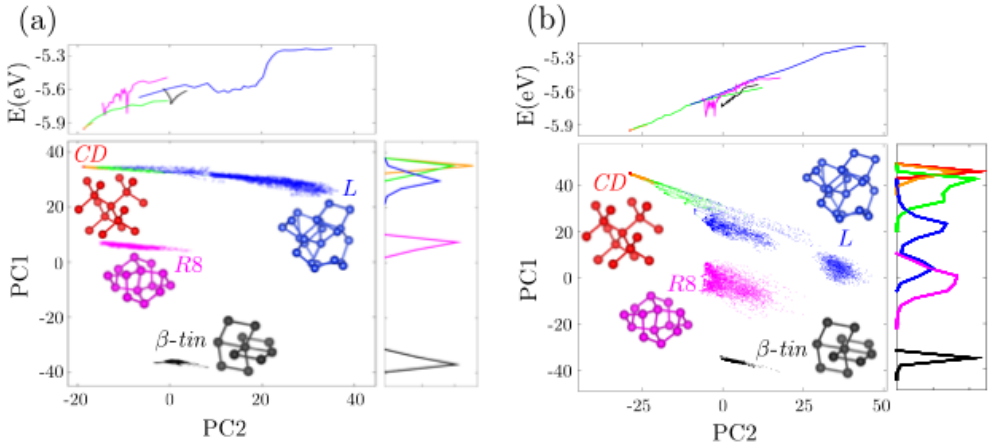


Figure 3.2: (a) PCA of the symmetry function representation. (b) PCA of the hidden layer representation. The central panels are the representations of Si structures projected onto the first two principal components, PC1 and PC2. The side panels are: Right - a histogram of structures as a function of their PC1 value to facilitate visualization of overlap; Top - average potential energy per atom (in eV) of structures plotted against PC2, calculated using a fixed bin size of 50 structures.

neighbor shell are presented in three different colors that identify the degree of crystalline order: green if the four nearest neighbors of the Si atom form a distorted tetrahedron[69,81,132], orange if they have perfect tetrahedral order in the nearest neighbor shell, and red if they have perfect tetrahedral order in the first and second neighbor shells. In the two-dimensional projection of the symmetry function space, the different phases are already separated in clusters, Fig. 3.2(b). Although there are hundreds of red points on this plot, all the perfectly ordered *CD* structures are contained within a very narrow region of (PC1,PC2) values. PC1 can be identified as the phase information, although it is not able to distinguish between different levels of order in the *CD* structures very well. *β-tin* (magenta) structures form an isolated island in PC1. In the two dimensional projection of the hidden layer representation Fig. 3.2(b), PC1 distinguishes between different phases with higher precision. The level of order present in *CD* structures, between red, green and blue is well represented in PC1. *L*

(blue) structures have two distinct regions, the higher temperature liquid-like regions overlap with the *R8* phase in PC1, whereas the lower temperature *L* has some overlap with the *CD* (green) phase. The *β-tin* phase is a 6-fold metallic structure whereas *R8*, *CD* and the *L* phases all represent different structures with various degrees of 4-fold tetrahedral coordination. The *R8* phase can transform to the *CD* phase at zero pressure whereas a pressure of $\sim$ 11.7 GPa is required to transform the *CD* phase to the *β-tin* phase[133]. Interestingly, this is captured in the hidden layer representation, since the data collected at zero pressure presents *β-tin* as a separate island in PC1, whereas the symmetry function representation does not present this distinction. PC2 is related to the potential energy of the structures. In Fig. 3.2(a) we show the potential energy, which is correlated with the PC2 value for different phases. The dependence is not as strong for the *β-tin* phase, as the energy of the *β-tin* structures seems to go down and then up as a function of PC2. In the hidden layer representation the correlation is stronger [Fig. 3.2(b)]: the energy of structures in each phase is a linear function of PC2. It is interesting that the last layer of the NN projects the hidden layer representation onto PC2 to calculate the energy of a structure.

The PCA showed us that, through non-linear transformations, the NN represents the original data by emphasizing the structural phase and energy information. Not surprisingly, that the PC1 of symmetry functions is related to the phase of a structure, since phase is the main source of variance in the structure of atomic neighborhoods. It is interesting that the NN makes that relation stronger, by making the PC1 of the hidden layer describe phase information more precisely. Since specific information on the various phases of the Si structures were not supplied to the NN or used at any step of the learning process, as, for instance, in the definition of the symmetry functions, it is quite interesting that the NN learned to emphasize these key features

independently in order to represent correctly the energy landscape. Overall, Fig. 3.2 shows that the symmetry function representation is a good representation for expressing the potential energy, with PC1 and PC2 being the physically relevant quantities. The NN does not change the character of the symmetry function representation, but makes the relations stronger and more precise, which produces the almost *ab initio* level prediction of the total energy of each structure.

To gain a more detailed understanding of the representations, especially the non-linear and local properties of the data, we use the non-linear dimensionality technique t-SNE (t-Distributed Stochastic Neighbor Embedding)[134]. Unlike PCA, t-SNE focuses on the local topology of points in high dimensional space to project data onto a lower dimensional space, typically of size 2 or 3, which captures how the individual atom neighborhoods are represented within the NN, in relation to similar neighborhoods around it.

In order to investigate a single Si atom's interaction with its neighbors, we study the representation of local neighborhoods as we step through the NN, beginning with a Cartesian description and ending in the final hidden layer. Before the application of symmetry functions and the NN, the atoms are represented in Cartesian coordinates. An atom's neighborhood is represented by $3 \times N_a$ numbers, the coordinates of N_a neighbors relative to the central atom at the origin. The symmetry functions are calculated with a cutoff of 6 Å, which on average includes 43 neighbors. We apply t-SNE to local neighborhoods in the Cartesian space, which is $(3 \times 43) = 129$-dimensional. We sort the coordinates of neighbors in order of their distance, that is, the n^{th} set of three numbers denote the position of the n^{th} nearest neighbor. This representation is invariant to translations and permutations of atoms, but not to rotations. We apply t-SNE to 50,000 randomly chosen local neighborhoods in the Cartesian repre-

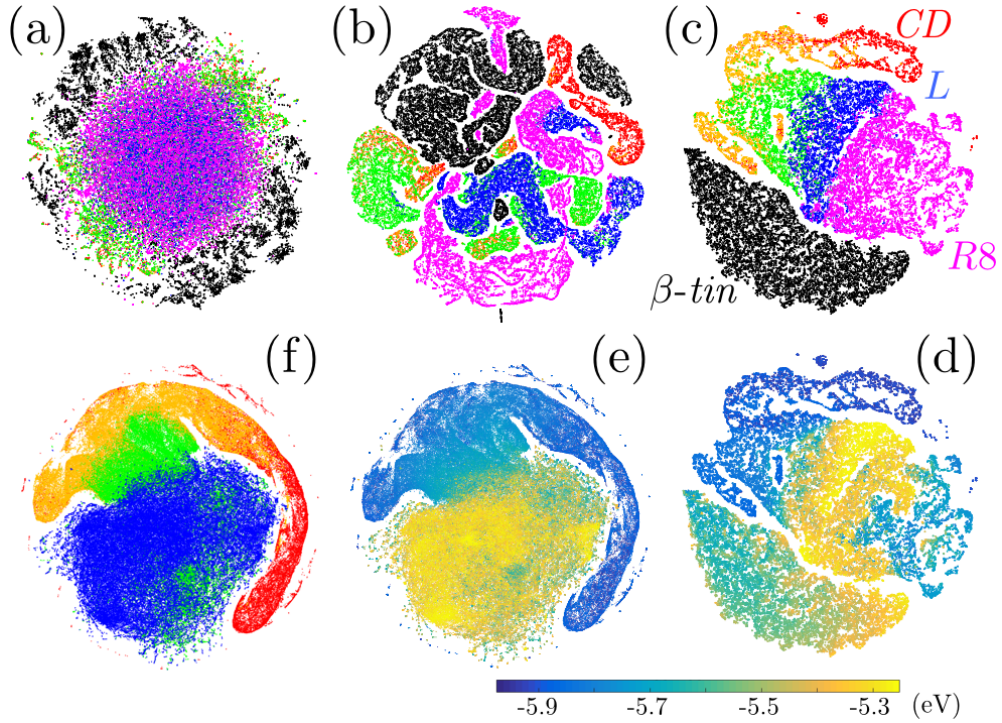


Figure 3.3: Representations of the Si local neighborhoods by t-SNE: (a), (b), (c) are maps of the Cartesian, symmetry function, and hidden layer representations, respectively. Each map has 50,000 neighborhoods. The color of each phase is the same as in Fig. 3.2. (d): NN hidden layer colored according to the energy contribution of neighborhoods. (e), (f): Hidden layer representation of the NN trained only on the *CD* and *L* phases.

sentation, which we show in Fig. 3.3(a). There is a clear separation between the *β-tin* phase and the other phases. The *R8* and *L* phases are completely mixed, and the *CD* phase is scattered irregularly.

The next step of the NN pipeline is the calculation of the symmetry functions. We use t-SNE to map the 28-dimensional symmetry function space to 2 dimensions, as shown in Fig. 3.3(b). The symmetry function representation is invariant to rotations, in addition to the symmetry invariances of the Cartesian representation. Since radial (angular) symmetry functions are calculated from two-body (three-body) type interactions, they are more convenient for describing the contributions of atomic in-

teractions to the energy. This is apparent in Fig. 3.3(b), where neighborhoods in the same phase are clustered together. *CD* phases at different tetrahedral levels of order are mixed together. Some of the local neighborhoods from the *R8* phase are mixed with some of the *L* neighborhoods. The clustering of similar neighborhoods is the reason that many existing ML algorithms, even linear models (in the case of metallic atoms)[67], have successfully modeled the energy landscape of atomistic systems using symmetry functions. In contrast, it would be much harder to perform ML techniques on the Cartesian representation directly.

Next, we apply t-SNE to the hidden layer representation [Fig. 3.3(c)]. The *β-tin* phase is completely separated, as it was in the first two dimensions of the PCA. *L* neighborhoods are connected to both *R8* and *CD*. Neighborhoods in the *CD* phase farther from the *L* phase have better tetrahedral order, in a connecting, neck-like structure color-coded from green to orange to red to indicate increasing order. This set of structures represent the region of phase transition from the *L* to the perfect *CD* order.

3.5 Discussion

In the remaining discussion we offer arguments of how this analysis can actually lead to interesting physical insight on possible phase transformations in Si. As a first step in this direction, in Fig. 3.3(d), we present the same map with each point colored not by its phase label, but according to the energy contribution of that local neighborhood. This energy is not defined within DFT, but it is defined as the output of each atom's individual NN. This plot shows how the neighborhoods within a phase are distributed: the higher energy neighborhoods of the *R8* phase are closer to the *L* region; the *R8* and *CD* phases are both connected to the *L* phase, but the energy is lower for

neighborhoods that are farther from the L region; finally, the β-*tin* neighborhoods have generally higher energies, with those on the right side at a higher potential energy compared to those on the left side of the corresponding phase region.

To see the structural transition from the L phase to the *CD* crystal in more detail, we train a separate NN on just these two phases. We apply t-SNE on the representation of this NN [Fig. 3.3(f)] which allows us to see the structure of the hidden layer representation on these two phases in more detail. There is a path from liquid-like Si neighborhood to *CD* crystal through a smooth transition of increasing order. This transition is also apparent in Fig. 3.3(e). The higher energy L neighborhoods smoothly transform into the lowest energy *CD* neighborhoods. To offer a visualization of this general shape in the original 60-dimensional representation space, we could describe it as a "sphere" of neighborhoods with no tetrahedral order connected by a long "neck" that extends from the "sphere" all the way to the perfect crystal. Including the potential energy as an additional dimension, morphs this representation into a landscape with a single funnel extending to the global minimum. It is well known that Si crystalizes to the *CD* phase easily during the cooling[135]. This can be explained by Si having a deep global minimum at the bottom of a single funnel that is easily reached[136,137].

A second insight gained by studying these maps relates to the transformation between crystalline phases of Si. For example, the fact that high external pressure is required for the transition from the *CD* crystalline phase to the β-*tin* crystalline phase is explained by the feature of the neighborhood map that the local neighborhoods corresponding to the β-*tin* structure lie on a separate manifold than those of the L and *CD* phases. A significant energy barrier (through the applied pressure) must be overcome to pass from one phase to the other, as has been established both experi-

mentally[138,139] and theoretically[140].

ML algorithms are gaining popularity in solid state physics and chemistry[70,71,120,141–144]. Due to increasing computational resources, the wider availability of scientific data and advanced ML algorithms, data science approaches can prove very effective for studying complex atomistic problems. Instead of merely using ML as a "black box" approximation tool, we have provided examples of how it can be used to "learn" scientific concepts from computational and experimental data. In a different context, we used ML representations combined with theoretical models to gain insights about the relationship between structural and dynamical properties of atomistic systems[70–72]. We have also shown that the application of the ML algorithm can be interpreted as a set of physically meaningful transformations on data. We expect that these notions can be extended to address issues related to more complex materials.

4

Identifying structural flow defects in disordered solids using machine learning methods

4.1 Overview

We use machine learning methods on local structure to identify flow defects – or regions susceptible to rearrangement – in jammed and glassy systems. We apply this method successfully to two disparate systems: a two dimensional experimental realization of a granular pillar under compression, and a Lennard-Jones glass in both two

and three dimensions above and below its glass transition temperature. We also identify characteristics of flow defects that differentiate them from the rest of the sample. Our results show it is possible to discern subtle structural features responsible for heterogeneous dynamics observed across a broad range of disordered materials.

4.2 Introduction

All solids flow at high enough applied stress and melt at high enough temperature. Crystalline solids flow[145] and premelt[146] via localized particle rearrangements that occur preferentially at structural defects known as dislocations. The population of dislocations therefore controls both how crystalline solids flow and how they melt. In disordered solids, it has long been hypothesized that localized particle rearrangements[147] induced by stress or temperature also occur at localized flow defects[148–150]. Like dislocations in crystals[151], flow defects in disordered solids are particularly effective in scattering sound waves, so analyses of the low-frequency vibrational modes[152] have been used successfully to demonstrate the existence of localized flow defects [43,151,153–160]. However, all attempts to identify flow defects[161,162] directly from the structure, without using knowledge of the inter-particle interactions, have failed[161,162]. Likewise, in supercooled liquids, purely structural measures correlate only weakly with kinetic heterogeneities[160], although correlations between structure and dynamics have been established indirectly[163–167].

Here we introduce a novel application of machine learning (ML) methods to identify "soft" particles that are susceptible to rearrangement or, equivalently, that belong to flow defects, from the local structural geometry alone. We apply the method to two very different systems–an experimental frictional granular packing under uniaxial compression and a model thermal Lennard-Jones glass in both two and three

dimensions. The analysis of granular packing shows that our method succeeds even when previous methods based on vibrational modes[43] are inapplicable. The results for Lennard-Jones systems show that the correlation between structure and irreversible rearrangements does not degrade with increasing temperature, even above the dynamical glass transition, and is equally strong in two and three dimensions. Finally, we exploit the method to discover which structural properties distinguish soft particles from the rest of the system, and to understand why previous attempts to identify them by structural analysis have failed.

4.3 Methods

Physically-motivated quantities such as free volume or bond orientational order correlate with flow defects[43], but are insufficient to identify them *a priori*. We introduce a large set of quantities that are each less descriptive, but when used as a group provide a more complete and unbiased description of local structure. These quantities have been used to represent the potential energy landscape of complex materials from quantum mechanical calculations[65]. For a system composed of multiple species of particles, we define two families of structure functions for each particle i,

$$G_Y^X(i;\mu) = \sum_j e^{-(R_{ij}-\mu)^2/L^2} \tag{4.1}$$

$$\Psi_{YZ}^X(i;\xi,\lambda,\zeta) = \sum_j \sum_k e^{-(R_{ij}^2+R_{ik}^2+R_{jk}^2)/\xi^2}(1+\lambda\cos\theta_{ijk})^\zeta \tag{4.2}$$

where R_{ij} is the distance between particles i and j; θ_{ijk} is the angle between particles i, j and k; L, μ, ξ, λ, and ζ are constants; X,Y,Z are labels that identify the different species of particles in the system, with the correspondence $i \leftrightarrow X$, $j \leftrightarrow Y$, $k \leftrightarrow Z$. By

using many different values of the constants μ, ξ, λ, and ζ we generate many structure functions in each family that characterize different aspects of a particle's local configuration. The first family of structure functions G characterizes radial density properties of the neighborhood, while the second family, Ψ, characterizes bond orientation properties. The sums are taken over particle pairs whose distance is within a large cutoff R_c^S. Our results are qualitatively insensitive to the choice of R_c^S as long as it includes several neighbor shells.

In this study we use a range of parameters μ, ξ, ζ, and λ to represent local structures. For a particle of species X (where X is A or B) there will be two radial functions G_A^X and G_B^X for each value of μ and three angular functions Ψ_{AA}^X, Ψ_{AB}^X, and Ψ_{BB}^X for each triple (ξ, λ, ζ). We take μ between $0.3\sigma_{AA}$ and $5.0\sigma_{AA}$ in increments of $0.1\sigma_{AA}$, which gives a total of 94 radial structure functions per particle. For the angular structure functions, the parameters ξ, ζ, and λ are shown in Table 4.1. We use a total of 66 angular structure functions per particle. The classification accuracy is not very sensitive to the number of structure functions, as long as enough structure functions are included to describe the local neighborhood of particles.

Having characterized local structure through $G_Y^X(i;\mu)$ and $\Psi_{YZ}^X(i;\xi,\lambda,\zeta)$, we introduce a method to infer from this information the location of flow defects in disordered solids. Generically, we begin with a set of N particles to be classified as "soft" or "hard." Each particle is described by a set of M variables derived from the two families of functions G_Y^X and Ψ_{YZ}^X by varying the constants μ, ξ, λ, and ζ (here $M = 160$); this is represented by the set of vectors $\{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_N\}$, where $\boldsymbol{F}_i$ constitutes an embedding of the local environment of a particle i, constructed at a time t_i, in $\mathbb{R}^M$. We select at random a subset of n of these particles (the "training set") and categorize them *a posteriori* as being soft if they rearrange (the details of which will be dis-

Structure function parameters for Ψ_{YZ}			
	ξ (σ_{AA})	ζ	λ
1	14.633	1	-1
2	14.633	1	1
3	14.638	2	-1
4	14.638	2	1
5	2.554	1	-1
6	2.554	1	1
7	2.554	2	-1
8	2.554	2	1
9	1.648	1	1
10	1.648	2	1
11	1.204	1	1
12	1.204	2	1
13	1.204	4	1
14	1.204	16	1
15	0.933	1	1
16	0.933	2	1
17	0.933	4	1
18	0.933	16	1
19	0.695	1	1
20	0.695	2	1
21	0.695	4	1
22	0.695	16	1

Table 4.1: Angular structure function parameters.

cussed below) between time t_i, when the structure is characterized, and time $t_i + \Delta t$. Otherwise the particles are labeled as hard.

The next step is to use the particles in the training set, already classified as soft or hard, to construct a scheme to identify other particles as soft or hard. We use the support vector machine (SVM) method[168], which constructs a hyperplane in $\mathbb{R}^M$ that best separates soft particles from hard ones. Once this hyperplane has been established for the training set, the rest of the particles (and any particles from similar systems) may be classified according to whether their local structures place them on soft or hard sides of the hyperplane. Generically, no exact separation is possible so the SVM method is adapted to penalize particles whose classification is incorrect; the

degree of penalty is controlled by a parameter C where larger values of C allow for fewer incorrect classifications. We find that the quality of our classifications is insensitive to C for $C > 0.1$(see Section 4.3.1). The SVM algorithm was implemented using the LIBSVM package[169].

To identify rearrangements we calculate, for each particle i, the widely-used quantity[149]

$$D^2_{\min}(i) \equiv \min_{\Lambda} \left\{ \frac{1}{z} \sum_j (R_{ij}(t + \Delta t) - \Lambda R_{ij}(t))^2 \right\}, \tag{4.3}$$

which characterizes the magnitude of non-affine displacement during a time interval Δt. Here the sum runs over neighbors j within a distance of R_c^D of particle i, R_{ij} is the center-center displacement between particles i and j, z is the number of neighbors within R_c^D. The quantity is minimized over choices of the local strain tensor Λ. We find $D^2_{\min}$ is insensitive to the choice of R_c^D and Δt so long as R_c^D is large enough to capture the particles local neighborhood and Δt is longer than the ballistic timescale. A particle is said to have undergone a rearrangement if $D^2_{\min} \geq D^2_{\min,0}$. We choose $D^2_{\min,0}$ such that approximately 0.15% of the particles from each species in each system has gone through a rearrangement although the results depend little on the specific choice of $D^2_{\min,0}$, as explained below.

4.3.1 Optimization of $D^2_{\min,0}$, svm parameters, and training set size

To justify our choice of $D^2_{\min,0}$, we consider how the accuracy of the classification depends on $D^2_{\min,0}$ at $T = 0.4$ for the $d = 2$ LJ system. Fig. 4.2 shows that the accuracy increases with $D^2_{\min,0}$ until about $D^2_{\min,0} = 0.3\sigma^2_{AA}$, after which it appears to level off. This is consistent with our choice of $D^2_{\min,0} = 0.28\sigma^2_{AA}$, which is also consistent with the $D^2_{\min}$ threshold used to define rearrangements in Ref.[159]. This

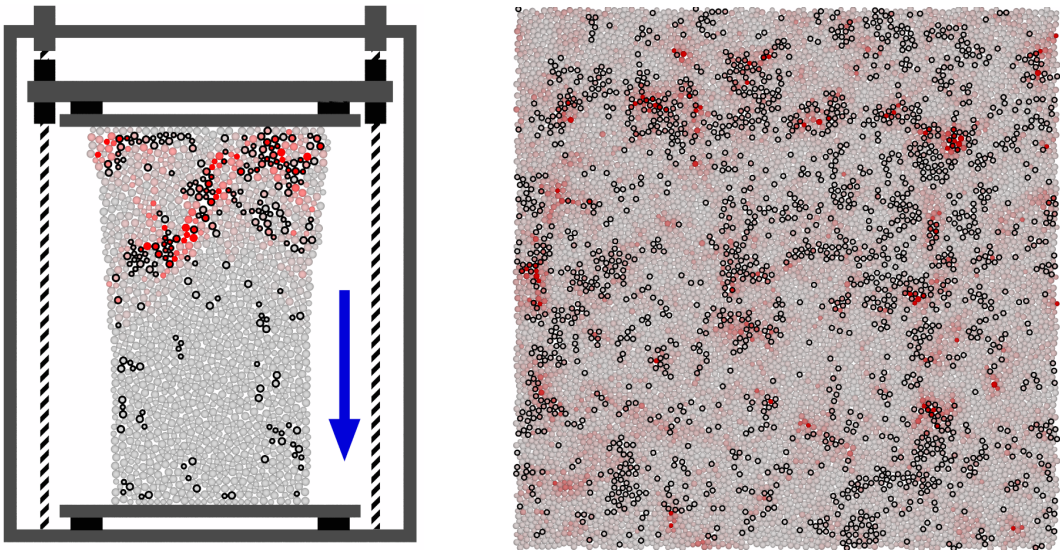


Figure 4.1: (color online) Snapshot configurations of the two systems studied. Particles are colored gray to red according to their D^2_{min} value. Particles identified as soft by the SVM are outlined in black. (a) A snapshot of the pillar system. Compression occurs in the direction indicated. (b) A snapshot of the $d = 2$ sheared, thermal Lennard-Jones system.

threshold value of $D^2_{\text{min},0} = 0.28\sigma^2_{AA}$ corresponds to a non-affine displacement of order $\sqrt{D^2_{\text{min},0}} = 0.53\sigma_{AA}$. The $D^2_{\text{min},0}$ for other temperatures were scaled linearly with temperature, to give $D^2_{\text{min},0}(T) \equiv 0.7T\sigma^2_{AA}$.

The training set is given by $\{(\boldsymbol{F}_1, y_1), ..., (\boldsymbol{F}_N, y_N)\}$, where $\boldsymbol{F}_i = \left\{F_i^1, ..., F_i^M\right\}$ are the values of the structure functions that describe the local neighborhood of particle i. $y_i = 1$ if the neighborhood i rearranges within time Δt, and $y_i = -1$ otherwise. We aim to find a hyperplane $\boldsymbol{w} \cdot \boldsymbol{F} - b = 0$ that separates the points with different y_i.

The SVM algorithm was implemented using the LIBSVM package[169]. Due to the stochastic nature of rearrangements (i.e. not all soft particles rearrange at every time interval) there does not exist a hyperplane that perfectly separates the two different classes. For this reason, a penalty parameter C is introduced, and we find the optimal

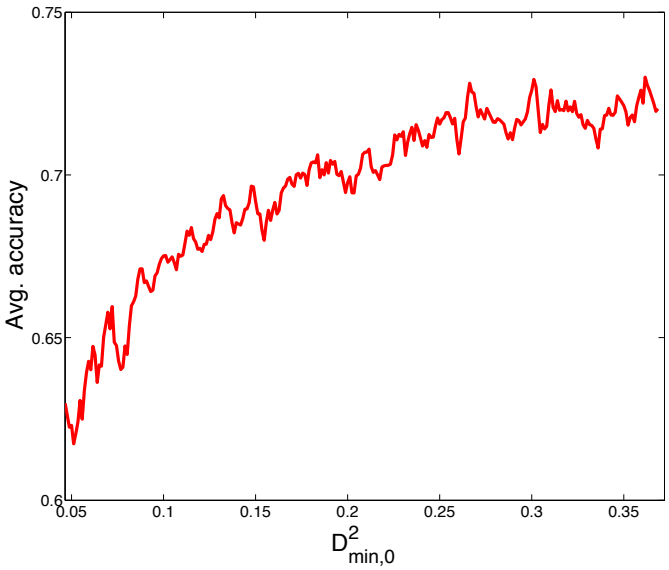


Figure 4.2: Identification accuracy over all hard and soft particles in the test set, as a function of $D^2_{\text{min},0}$ in units of σ_{AA}.

hyperplane equation by minimizing

$$\frac{1}{2}\boldsymbol{w}^T \cdot \boldsymbol{w} + C\sum_{i=1}^{N}\xi_i, \tag{4.4}$$

subject to the constraint that $y_i \cdot \left(\boldsymbol{w}^T \cdot \phi(\boldsymbol{F}_i) + b\right) \geq 1 - \xi_i$ and $\xi_i \geq 0$. When solved in the dual form, $\phi(\boldsymbol{F}_i)$ is not explicitly needed. Instead, the kernel function defined by $K(\boldsymbol{F}_i, \boldsymbol{F}_j) = \phi(\boldsymbol{F}_i) \cdot \phi(\boldsymbol{F}_j)$ is sufficient. In this study, we tried the linear ($K(\boldsymbol{F}_i, \boldsymbol{F}_j) = \boldsymbol{F}_i \cdot \boldsymbol{F}_j$) and radial basis function kernels ($K(\boldsymbol{F}_i, \boldsymbol{F}_j) = \exp\left\{-\gamma||\boldsymbol{F}_i - \boldsymbol{F}_j||^2\right\}$). The tunable parameters for both the linear and the RBF kernels were optimized through cross-validation in order to obtain the best generalizability. Through exhaustive search, we found that for the optimal parameters γ and C, the linear kernel performed as well as the RBF kernel. For this reason we use the linear kernel for all of the results presented here. For the linear kernel, we find that the results are not very sensitive to

the tuning parameter for sufficiently large C ($C > 0.1$), for both the LJ glass and pillar systems. In all cases, we used $C = 1$ in our calculations. Fig. 4.3 shows that the fraction of misclassified soft and hard particles decreases with increasing C, as expected.

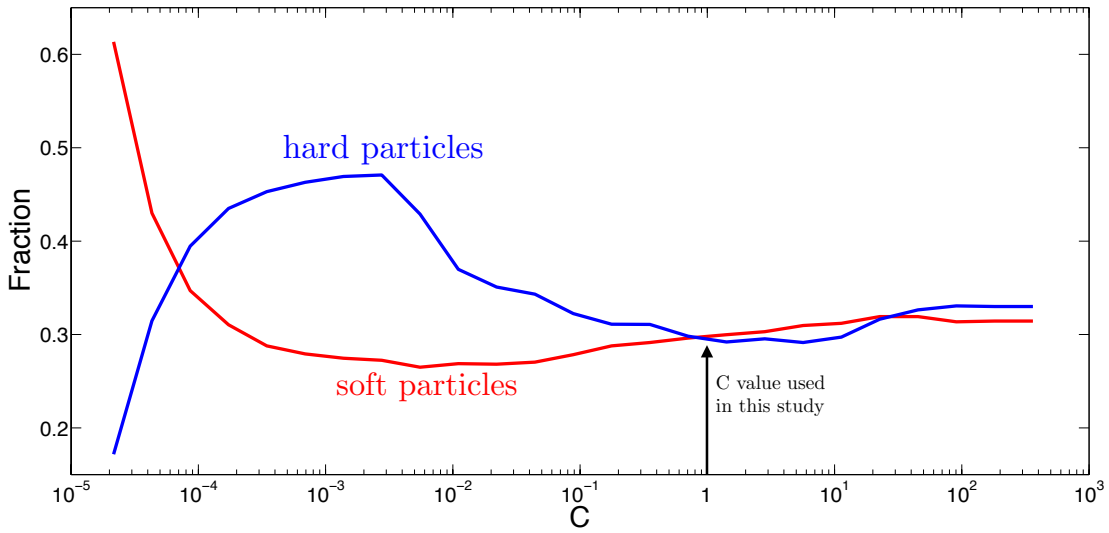


Figure 4.3: Fraction of misclassified soft and hard particles at different C values.

To determine the size of the training set needed to achieve a good classifier, we trained SVMs on sets of different sizes and applied the SVM to a test set of 10,000 points, for the $d = 3$ LJ glass. Fig. 4.4 shows that for training sets with fewer than 12,000 points, the accuracy of classification on the training set is higher than for the test set, which indicates overfitting. However, for training sets with more than 12,000 points, the test set accuracy is maximized. Furthermore, the test set accuracy is equal to the training set accuracy, which is optimal.

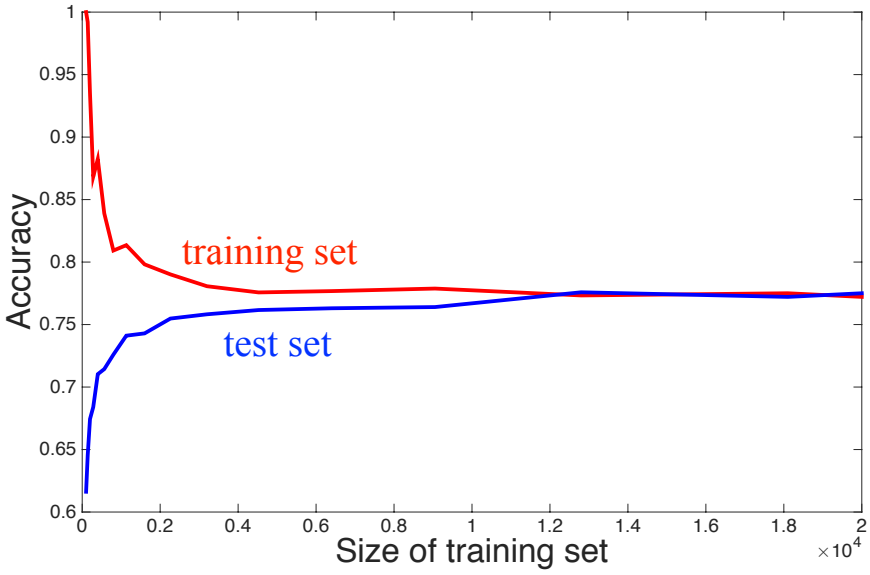


Figure 4.4: Classification accuracy on the training and the test set as a function of training set size.

4.4 Results

We first test our approach on an experimental system of two-dimensional ($d = 2$) "pillars" of particles. A bidisperse pillar made up of grains (plastic cylinders resting upright on a horizontal substrate) is situated between two plates in a custom-built apparatus. The bottom plate is fixed and the top plate is driven into the pillar at a constant speed of $v_0 = 0.085$ mm/s. The pillars are composed of a bidisperse mixture of approximately 1500 rigid grains with size ratio 3:4 and the large particles having a radius of $d_{AA} = 0.3175$cm. These particles have elastic and frictional interactions with each other, as well as frictional interactions with the substrate, making the identification of flow defects using vibrational modes impossible. A camera is mounted above and captures images at 7 Hz throughout the compression.

We construct our training set from compression experiments performed on ten dif-

ferent pillars. We select 6,000 particles at random from the entire duration of the experiments that undergo a rearrangement in the next 1.43 seconds and an equal number of particles that do not. To identify rearrangements we calculate D^2_{min} with $R^D_c = 1.5 \cdot d_{AA}$ and $D^2_{\text{min},0} = 0.25 d^2_{AA}$. Compression of the mechanical pillar from the top only affects particles above a certain "front" that starts at the top and advances towards the bottom with time. Our training set contains only particles within this activated front. Particles in a horizontal slice between y and $y + \delta y$ are said to be within the activated front if the average speed of particles in the slice exceeds $v_{\text{thresh}} \sim 0.04$ mm/s.

As a second test, we apply our approach to a model glass in both $d = 2$ and $d = 3$. We study a 65:35 binary Lennard-Jones (LJ) mixture with $\sigma_{AA} = 1.0$, $\sigma_{AB} = 0.88$, $\sigma_{BB} = 0.8$, $\epsilon_{AA} = 1.0$, $\epsilon_{AB} = 1.5$, and $\epsilon_{BB} = 0.5$[170]. The LJ potential is cut off at $2.5\sigma_{AA}$ and smoothed so that both first and second derivatives go continuously to zero at the cutoff. The natural units for the simulation are σ_{AA} for distances, ϵ_{AA} for energies, and $\tau = \sqrt{m\sigma^2_{AA}/\epsilon_{AA}}$ for times. We perform molecular dynamics simulations using LAMMPS[171] with a time step of $5 \times 10^{-3}\tau$ at density $\rho = 1.2$, using a Nosé-Hoover thermostat with a time constant of τ. Here, temperature is in units of ϵ_{AA}/k_B, where k_B is the Boltzmann constant. In $d = 2$, we consider a system of 10,000 particles at temperatures $T = 0.1, 0.2, 0.3$, and 0.4 at a single strain rate, $\dot{\gamma} = 10^{-4}/\tau$. In all cases data was collected after allowing the system to reach steady state by shearing up to 20% strain. In $d = 3$, we use a collection of 30,000 particles at temperatures $T = 0.4, 0.5, 0.6$ with no strain. The quiescent system has a glass transition at $T_G = 0.33$ in $d = 2$ and $T_G = 0.58$ in $d = 3$[170]. Therefore, in both dimensions we study the system both above and below its glass transition temperature.

At each temperature we construct training sets of 6,000 and 10,000 particles, in

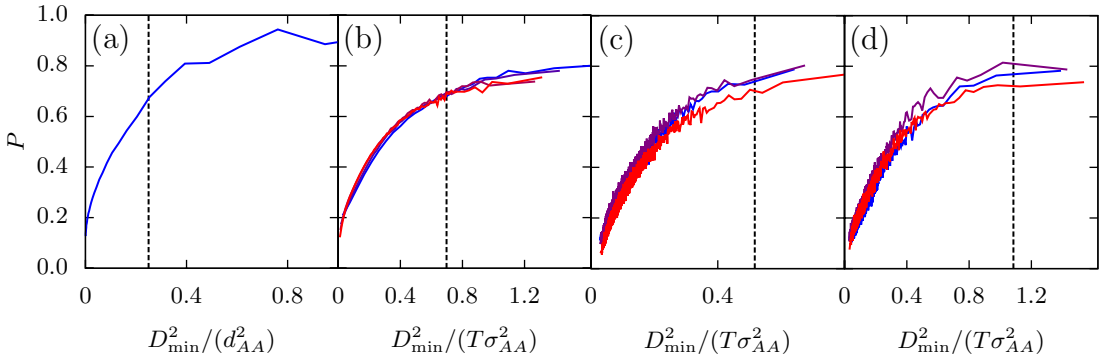


Figure 4.5: (color online) Probability that a particle of a given $D^2_{\rm min}$ value is soft. The vertical dashed lines are corresponding $D^2_{\rm min,0}$ values. (a) The result for the pillar system, where d_{AA} refers to the large grain diameter (since this is a granular system with macroscopic grains, thermal fluctuations are negligible). (b) The result of using an SVM trained at a temperature T ($T = 0.1, 0.2, 0.3$ and 0.4 shown in different colors) to classify data at the same temperature for the $d = 2$ LJ glass. (c,d) Results for species A and B, respectively, for the $d = 3$ system at $T = 0.4, 0.5$ and 0.6.

$d = 2$ and 3, respectively, selected at random from the entire run that undergo a rearrangement in the following $\Delta t = 2\tau$ units of time, and an equal number of particles that do not undergo a rearrangement. To identify rearrangements, we calculate $D^2_{\rm min}$ with $R^D_c = 2.5\sigma_{AA}$ to be the same as the range of the truncated LJ potential and $\Delta t = 2\tau$. In $d = 3$ the $D^2_{\rm min}$ distributions of the species A and species B particles differ significantly and so are treated separately throughout the analysis.

We now test our classifying hyperplane on the three systems outlined above. For each system we construct a test set of particles that were not used in training the SVM; for the pillar this test set consists of 100,000 particle environments from ten additional compression experiments. For the $d = 2$ and $d = 3$ LJ glasses we use 2×10^7 and 75×10^7 unseen particle environments respectively. Fig. 4.6 (a)-(d) shows the probability, $P(D^2_{\rm min})$, that a particle with an observed value of $D^2_{\rm min}$ was identified as soft by our classification *a priori* for each system. Fig 4.6 (c)-(d) treat the particles of species A and species B separately for the $d = 3$ LJ glass. In all cases we

see that $P(D^2_{\rm min})$ rises with increasing plastic activity, $D^2_{\rm min}$. This implies that the particles identified as soft by the SVM are more likely to be involved in plastic flow. For the granular pillar, 21% of the particles are classified as soft, and these particles capture 80% of the rearrangements. For the $d = 2$ and $d = 3$ LJ systems, these numbers are 26% and 73%, and 24% and 72%, respectively, at the highest temperatures studied. Thus, we consistently find rearrangements to occur at soft particles about 3-4 times more frequently than if the soft particles were randomly chosen. Finally, Fig. 4.6(b,c,d) show that $P(D^2_{\rm min})$ collapses for different T both for 2D and 3D systems, when $D^2_{\rm min}$ is scaled by $T\sigma^2_{AA}$; this scaling arises since for particles not undergoing rearrangement, $D^2_{\rm min} \sim \langle v^2 \rangle \sim T$ by the equipartition theorem[159].

4.4.1 Prediction accuracy at different temperatures

Remarkably, our ability to identify soft particles does not decrease with increasing temperature or dimension. For the $d = 2$ LJ system over the same temperature range, the accuracy of the vibrational mode method decreases by over 50%[159]. The key difference between the two methods is that we construct local environments from the actual particle positions in snapshots of the thermal system, while soft particles from vibrational modes are extracted from particle positions in the inherent structures obtained by quenching to $T = 0$.

The probability that a rearranging particle is soft is independent of temperature from $T = 0.1$ to $T = 0.4$. However, the corresponding probability decreases by more than a half when the soft particles are identified from the quasi localized vibrational modes (QLM method)[159]. Part of this difference arises the fact that QLM soft particles are extracted from the inherent structures of the system, whereas SVM soft particles are identified from thermal configurations.

To study the effect of this difference, we constructed a classification using the SVM method based on particle positions in the inherent structures (the same structures used to find the soft spots from the QLM method). We find that the ability to predict rearranging particles was reduced by 14% at higher temperatures (the probability that a rearranging particle is soft drops from 76% to 62%) as seen in Fig. 4.6. This implies that there is important structural information in snapshots of the thermal system that is lost in the inherent structures.

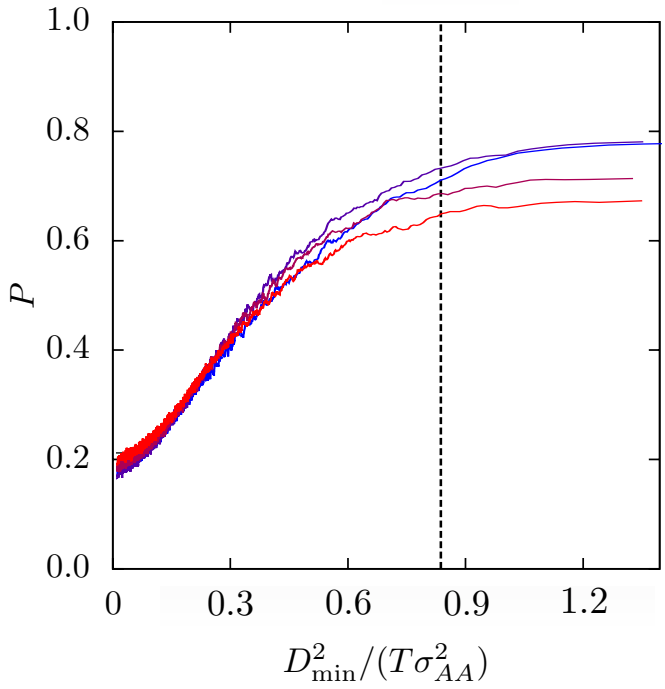


Figure 4.6: Probability that a particle of a given $D^2_{\rm min}$ value is soft. Here the soft particles are classified (in other words, the SVM is trained) using the inherent structures of the configuration. The vertical dashed line is $D^2_{\rm min,0}$. Here we show that as temperature increases the predictive power of the SVM degrades. This is contrasted with an SVM trained on thermal configurations, whose prediction accuracy does not depend on temperature.

4.5 Discussion

4.5.1 Generalizability of classification to different temperatures

It is interesting to ask if the local environments characterizing soft particles change as a function of temperature, or equivalently, whether the dividing hyperplane depends on temperature. We tested the temperature sensitivity of the hyperplane that divides soft from hard particles by using the hyperplane constructed at $T = 0.4$ to classify particles at the other three temperatures studied. The results of this procedure are shown in Fig. 4.7 (a). First, we notice that at all temperatures considered, P is once again an increasing function of $D^2_{\min}$. This implies that a hyperplane constructed at one temperature can be applied successfully over a range of temperatures. However, the correlation (and thus the accuracy of our method) decreases at temperatures farther away from the training temperature Fig. 4.7 (b). Thus, structural similarities in the environment of soft particles persist over a wide range of temperatures, but decrease with increasing temperature difference.

4.5.2 Physical interpretation

The hyperplane that divides soft from hard particles tells us which features of the local structural environment are important in distinguishing between the hard and soft particles. To illustrate this, we focus on the sheared $d = 2$ LJ glass at $T = 0.1$. The first set of structure functions, $G^X_Y(i;\mu)$ defined in Eq. (4.1), has a familiar physical interpretation in terms of the radial distribution function: $g_{XY}(r) = \lim_{L\to 0}\langle G^X_Y(i;r)\rangle/2\pi r$. This family of structure functions is essentially a parameterization of a particle's local contribution to $g(r)$. In Fig. 4.8 we show the approximations to various $g(r)$ constructed in this way, using $L = 0.1\sigma_{AA}$, for particles identified

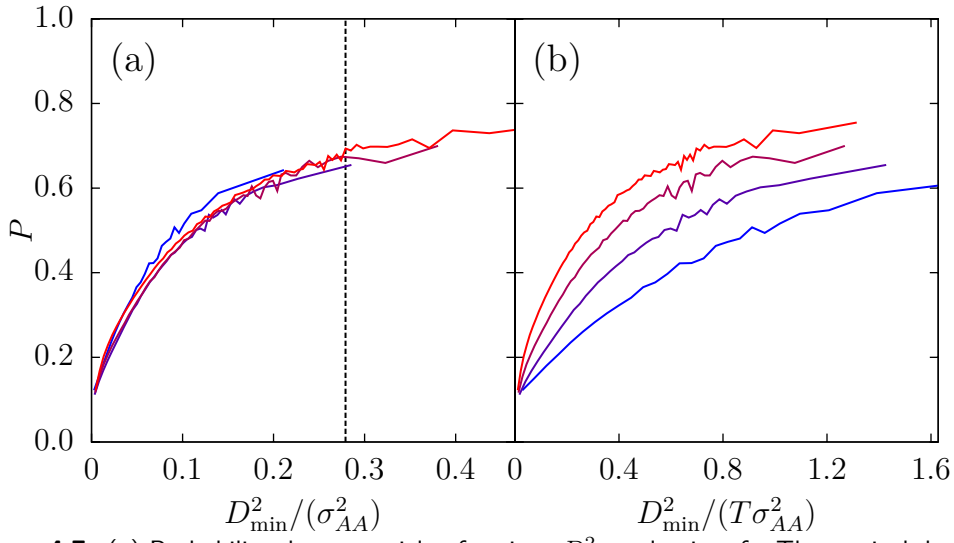


Figure 4.7: (a) Probability that a particle of a given $D^2_{\rm min}$ value is soft. The vertical dashed line is $D^2_{\rm min,0}$. Here we show the result of using an SVM trained at $T = 0.4$ to classify data at all temperatures. (b) Same as (a) but with $D^2_{\rm min}$ is scaled by corresponding T.

as hard (black lines) and soft (red lines).

In all cases, we see that soft particles feature slightly lower peaks and higher troughs in $g(r)$. To see whether this difference is sufficient to identify soft particles, we expand our analysis beyond average values of $G^X_Y(i;\mu)$, to the *distributions* of the different structure functions. In Fig. 4.9(a) we show the distribution of values of $G^B_A(i; r^{AB}_{\rm peak})$ where $r^{AB}_{\rm peak}$ is the location of the first peak of $g_{BA}(r)$, for soft particles (red) and hard particles (blue/green). While the distribution for soft particles features a single peak, that for hard particles is bimodal. This indicates the existence of (at least) two distinct populations of hard particles which we divide into two groups: one with $G^B_A(i; r^{AB}_{\rm peak})/r < 1/2$ (blue) that we will call H_0-type, and one with $G^B_A(i; r^{AB}_{\rm peak})/r > 1/2$ (green) that we will call H_1-type. Radial information therefore distinguishes between soft particles and H_1 hard particles but not between soft and H_0 hard particles.

We now consider the distribution of $\Psi^B_{AB}(i; 2.07\sigma_{AA}, 1, 2)$ for soft particles (red),

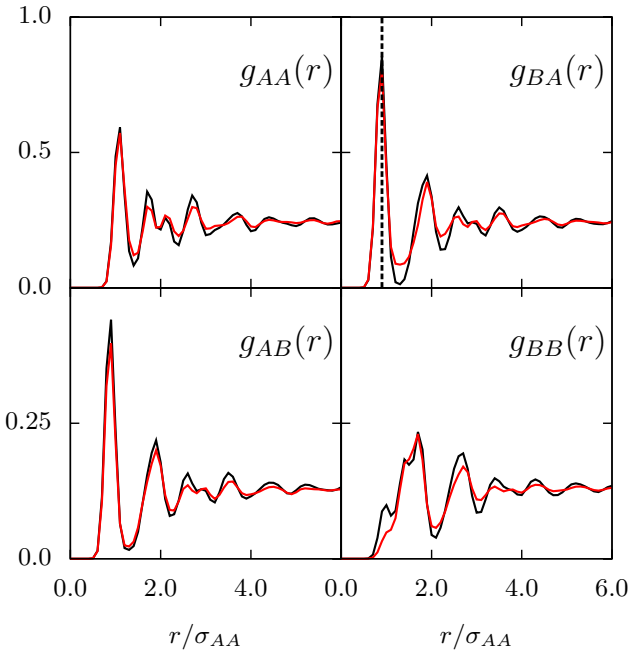


Figure 4.8: (color online) Radial distribution functions averaged over hard (black lines) or soft (red lines) particles. g_{AB} and g_{BA} of soft particles are not equal to each other since they refer to different kinds of regions: neighbors of soft particles from species A and neighbors of soft particles from species B, respectively.

H_0 hard particles (blue), and H_1 hard particles (green), shown in Fig. 3(c). Physically, $\Psi^B_{AB}(i; 2.07 r^{BA}_{\text{peak}}, 1, 2)$ is large when there are many pairs of neighbors of the central particle that lie within a distance ξ with small angles between them, such that one is of species A and one is of species B. The soft particles fall into a single category (one peak, red) while H_0-type and H_1-type hard particles–defined from radial information above–have very different distributions (blue and green peaks). Unlike before, here the soft particles and the H_0 hard particles have very different distributions while soft particles and H_1 hard particles have similar distributions. Bond-angle information therefore distinguishes between soft particles and H_0-type hard particles but not between soft particles and H_1-type hard particles. To fully distinguish be-

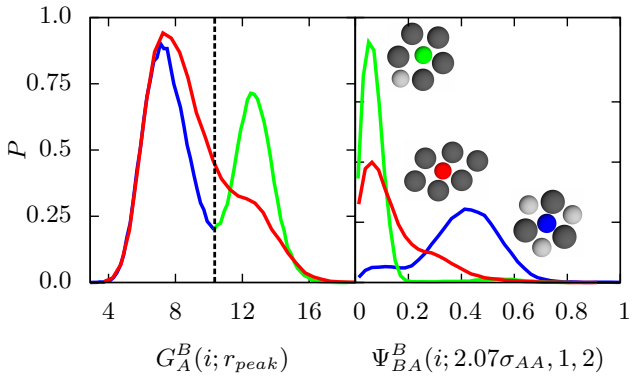


Figure 4.9: (color online) (a) Distribution of $G^A_B(i; r^{AB}_{\text{peak}})$, proportional to the gaussian weighted density at r^{AB}_{peak}, for soft (red) and hard (blue/green) particles. r^{AB}_{peak} corresponds to the first peak of g_{AB} or g_{BA}. (b) Distribution of $\Psi^B_{AB}(i; 2.07\sigma_{AA}, 1, 2)$, proportional to the density of neighbors with small bond angles near a particle i, for soft (red) and hard (blue/green) particles. The inset shows examples of configurations with corresponding radial and bond orientation properties, where dark (light) gray neighbors are of species A (B).

tween soft and hard particles, both radial and bond-angle information is needed. Soft particles have environments that–at a minimum–have fewer particles in their nearest neighbor shell *a*nd larger angles between adjacent particles.

4.6 Conclusion

In summary, we have presented a novel ML method for identifying flow defects in disordered solids. We note that we have focused on the short-time correlation of structure with particle rearrangements. However, our method should shed light on the connection between local structural evolution and the correlation of rearrangements in time and space[1] over longer time scales in glassforming liquids. We also note that we cannot predict the specific particles that will participate in rearrangements at a later time; rather, we identify a *population* of particles that are likely to rearrange. The latter quantity is more useful in thermal and/or sheared systems, since fluctuations lead to stochasticity in rearrangements.

Our method relies on local structure alone, and can be applied directly to snapshots of experimental systems, in contrast to previous methods[43]. Our approach also scales linearly with the number of particles, N, while vibrational mode approaches scale as N^3. The efficient identification of flow defects is critical to testing phenomenological approaches to plasticity based on flow defects[150,172–174]. Previous applications of machine learning methods in physics have focused on approximation[65,120,175] or on optimization and design tools[143,144,176,177]. Our approach shows that such methods–designed for detecting subtle correlations–can also be used directly to gain conceptual understanding not achieved with conventional approaches.

5

A structural approach to relaxation in glassy liquids

5.1 Overview

When a liquid freezes, a change in the local atomic structure marks the transition to the crystal. When a liquid is cooled to form a glass, however, no noticeable structural change marks the glass transition. Indeed, characteristic features of glassy dynamics that appear below an onset temperature, T_0,[44,45,178] are qualitatively captured by mean field theory[46,47,179], which assumes uniform local structure at all temperatures. Even studies of more realistic systems have found only weak correlations be-

tween structure and dynamics[43,160,161,180,181]. This raises the question: is structure important to glassy dynamics in three dimensions? Here, we answer this question affirmatively by using machine learning methods to identify a new field, "softness," which characterizes local structure and is strongly correlated with rearrangement dynamics. We find that the onset of glassy dynamics at T_0 is marked by the onset of correlations between softness (i.e. structure) and dynamics. Moreover, we use softness to construct a simple model of slow glassy relaxation that is in excellent agreement with our simulation results, showing that a theory of the evolution of softness in time would constitute a theory of glassy dynamics.

5.2 Introduction

To look for correlations between structure and dynamics, one typically tries to find a quantity that encapsulates the important physics, such as free volume, bond orientational order, locally preferred structure, etc. In contrast to this approach, we use a machine learning method designed to find a structural quantity that is strongly correlated with dynamics. In the previous chapter, we applied this approach to the simpler problem of classifying particles as being "soft" if they are likely to rearrange or "hard" otherwise. We describe a particle's local structural environment with $M = 166$ "structure functions"[65] that respect the overall isotropic symmetry of the system and include radial density and bond angle information. We then define an M-dimensional space, $\mathbb{R}^M$, with an orthogonal axis for each structure function. The local structural environment of a particle i is thus encoded as a point in M-dimensional space. We assemble a "training set" from molecular dynamics simulations consisting of equal numbers of "soft" particles that are about to rearrange and "hard" particles have not rearranged in a time τ_α preceding their structural characterization, and find the

best hyperplane separating the two groups using the support vector machines (SVM) method[168,169]. Finally, we define the *softness* S_i, of particle i as the shortest distance between its position in $\mathbb{R}^M$ and the hyperplane, where $S_i > 0$ if i lies on the soft side of the hyperplane and $S_i < 0$ otherwise.

We study a 10,000-particle 80:20 bidisperse Kob-Andersen Lennard-Jones glass[182] in $d = 3$ at different densities ρ and temperatures T above its dynamical glass transition temperature. All results here are for particles of species A only. However, the results are qualitatively the same for particles of both species. At each density we select a training set of $6,000$ particles, taken from a molecular dynamics trajectory at the lowest T studied, to construct a hyperplane in $\mathbb{R}^M$. We then use this hyperplane to calculate $S_i(t)$ for each particle i at each time t during an interval of $30,000\tau$ at each ρ and T.

5.3 Methods

5.3.1 System information.

We study a 10,000-particle Kob-Andersen model, a 80:20 binary LJ mixture[182] with parameters: $\sigma_{AA} = 1.0$, $\sigma_{AB} = 0.8$, $\sigma_{BB} = 0.88$, $\epsilon_{AA} = 1.0$, $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$, $m_A = m_B = 1$. Time is measured in units of $\tau = \sqrt{m_A \sigma_{AA}^2 / \epsilon_{AA}}$ and the Boltzmann constant is $k_B = 1$. We cut off the LJ potential at $2.5\sigma_{AA}$ and smooth the potential so that force varies continuously. This mixture has been characterized extensively. In particular, we compare our predictions to the measurements of the onset temperature in Keys *et al.*[1]. Simulations were done using LAMMPS[183] in an NVT ensemble with a Nosé-Hoover thermostat and a timestep of 0.0025τ. We output states every τ and quench them to their nearest inherent structure using a combination of conju-

Table 5.1: Number densities and temperatures studied. Each column contains the temperatures studied for a given number density ρ.

ρ	1.15	1.20	1.25	1.30
T	0.37	0.47	0.58	0.70
	0.42	0.51	0.61	0.75
	0.45	0.53	0.69	0.84
	0.52	0.56	0.76	0.92
		0.58	0.97	1.12
		0.70	0.97	

gate gradient and FIRE algorithms. Throughout this study we use inherent structure positions. However, qualitatively similar results can be obtained using time averaged positions. We study this system over the temperatures and number densities listed in Table 5.1.

5.3.2 Identifying rearrangements.

We identify rearrangements using the quantity p_{hop}. To define p_{hop} we first introduce a timescale $t_R = 10\tau$ that is chosen to be commensurate with the amount of time the system takes to complete a rearrangement. Two time intervals are defined as $A = [t - t_R/2, t]$ and $B = [t, t + t_R/2]$. An indicator function can then be written as,

$$p_{\text{hop}}(t) = \sqrt{\langle(\vec{r}_i - \langle\vec{r}_i\rangle_B)^2\rangle_A \langle(\vec{r}_i - \langle\vec{r}_i\rangle_A)^2\rangle_B} \tag{5.1}$$

where $\langle\rangle_A$ and $\langle\rangle_B$ are averages over the intervals A and B respectively.

When a particle is caged, p_{hop} is comparable in magnitude to the variance of a particle's position in its cage. If p_{hop} exceeds a threshold value we say that an event (a potential rearrangement) has occurred. We take this threshold to be 0.05, which is large compared to typical fluctuations of the particle in its cage. During an event,

p_{hop} rises above the threshold at some time t_1 (the start of the event), attains some maximum value p^*_{hop}, then drops below the threshold at a time t_2 (the end of the event). For each event we calculate the displacement of the particle during the event, $|\Delta r| = |r_t(t_2) - r_i(t_1)|$, as well as the duration of the event, $\Delta t = t_2 - t_1$. The latter quantity is related to the instanton time identified by Keys *et al.*[1]. Some typical events are shown in Fig. 5.1.

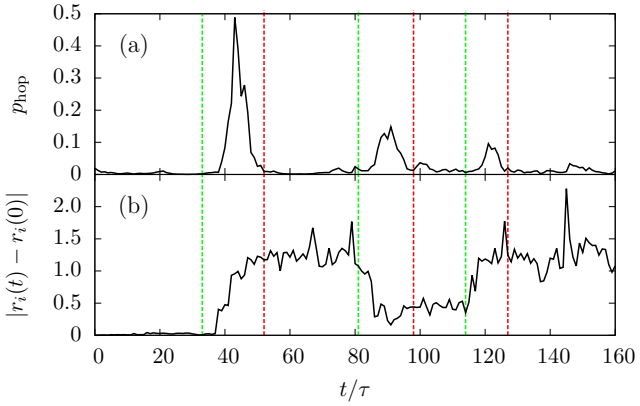


Figure 5.1: The typical trajectory of a particle at $T = 0.47$ and $\rho = 1.20$. (a) The values of p_{hop} over the course of the timeseries that contains three events. (b) The distance the particle has moved from its initial position over time. In green and red dashed lines indicate the beginning and end, respectively, of the events that we identify. Notice the clear separation of scales between events and the rest of the trajectory.

At each temperature, events corresponding to a particular value of p^*_{hop} are characterized by distributions of Δt and $|\Delta r|$. These are shown for several values of p^*_{hop} at $T = 0.47$ in Fig. 5.2. Generically, we see that the displacements and duration of events grow with increasing p^*_{hop}. We refer to the average displacement and timescale at a specific p_{hop} as $\langle|\Delta r|\rangle$ and $\langle\Delta t\rangle$ respectively.

The fact that events of a particular value of p^*_{hop} have a distribution of displacements is relevant when we compare the probability of rearrangement, $P_R(S)$, with the derivative of the overlap dq/dt. Since the former is defined in terms of p^*_{hop} and the

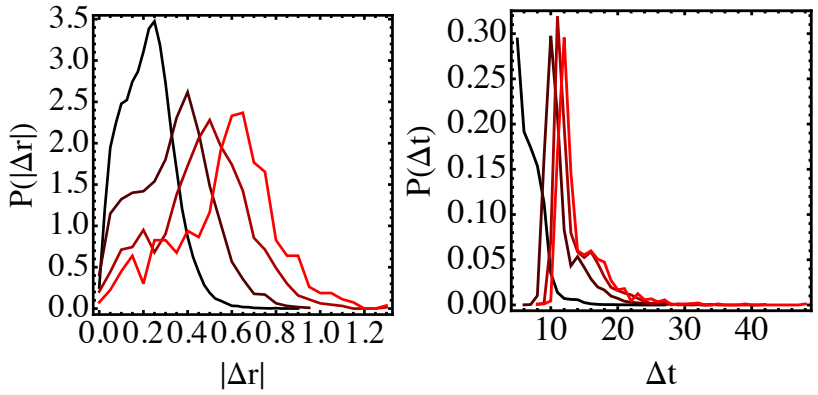


Figure 5.2: The characteristic size and timescale of events (potential rearrangements). (a) The distribution of displacements experienced by particles during rearrangements with $p^*_{\text{hop}} \approx 0.05$, 0.15, 0.25, and 0.35 at $T = 0.47$ from black to red respectively. (b) The distribution of durations of displacements observed at the same values of p^*_{hop} as in (a).

latter in terms of a distance a we must multiply $P_R(S)$ by the fraction of rearrangements in which particles move at least a distance a.

We now consider the dependence of $\langle|\Delta r|\rangle$ and $\langle\Delta t\rangle$ on temperature. In Fig. 5.3 (a) we show the dependence of $\langle|\Delta r|\rangle$ on both p^*_{hop} and temperature. We see that at all temperatures $\langle|\Delta r|\rangle$ increases nearly linearly with p^*_{hop}. This dependence allows us to use p^*_{hop} as a proxy for the scale of events. We note that the average size of events depends only weakly on temperature.

In Fig. 5.3 (b) we show how $\langle\Delta t\rangle$ depends on temperature and p^*_{hop}. We see that there is a relatively strong dependence of the duration of events on p^*_{hop} until p^*_{hop} reaches approximately 0.2, at which point (especially at low temperatures) it saturates at a time scale that we call $\tau_R(T)$. We note that $\tau_R(T)$ is the average duration of rearrangements at a temperature T (or the average instanton time of Keys *et al.*[1].) At low temperatures, $\tau_R(T)$ depends only weakly on temperature, but it increases with temperature. This time scale is important to the calculation of the evolution of the overlap parameter, as we will explain below.

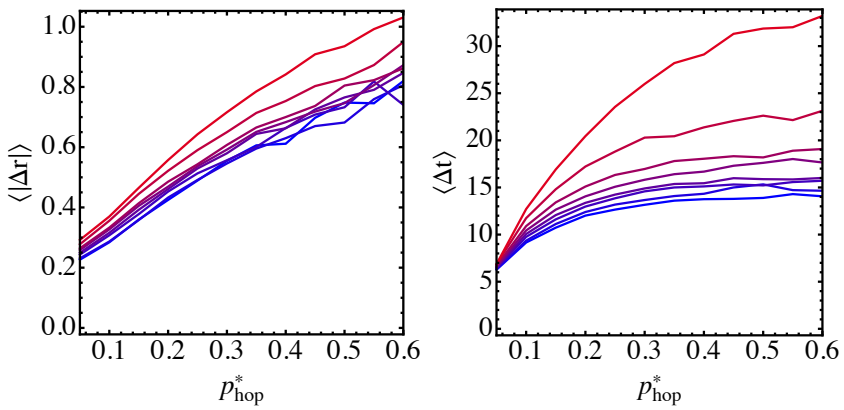


Figure 5.3: The average size and timescale of events. (a) The average size of events as a function of p^*_{hop} at temperatures $T = 0.45, 0.47, 0.51, 0.53, 0.56, 0.58, 0.63, 0.70$ from blue to red respectively. The average durations of events for the same temperatures as in (a).

Selection of the rearrangement cutoff

In this study, only those events with $p^*_{\text{hop}} > p_c = 0.2$ are considered to be rearrangements. Here we motivate this choice and explore how the Arrhenius processes that we identified in the text are affected by the cutoff. Fig. 5.4 shows the fraction of rearranging particles as a function of the cutoff p_c that are correctly identified as soft (*i.e.* that have $S > 0$) by the SVM at $T = 0.47$. We see that this prediction accuracy increases with p_c, saturating at $p_c \approx 0.2$, where the fraction of correctly identified particles reaches a plateau of about 90%. This is reminiscent of the behavior of the duration of an event, which also becomes independent of $p^*_{\text{hop}} \approx 0.2$ (Fig. 5.3 (b)). It therefore seems reasonable to consider as rearrangements only those events that exceed $p_c \approx 0.2$.

We now explore the relationship between the cutoff p_c and the energy and entropy scales extracted from the Arrhenius form. Fig. 5.5 (a) shows the energy scale as a function of softness for several different cutoffs. We see that the energy scales extracted from rearrangements at different cutoffs all have the same slope but experi-

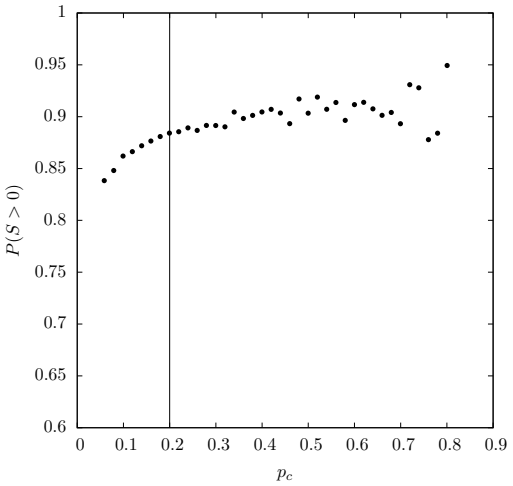


Figure 5.4: The fraction of rearranging particles that were correctly identified as soft as a function of p_c at $T = 0.47$ and $\rho = 1.20$. A line guiding the eye is drawn at $p^*_{\mathsf{hop}} = 0.2$.

ence a shift that is cutoff dependent. By contrast if we consider Fig. 5.5 (b) we notice that the entropy scale is virtually independent of softness in both its slope and shift. The fact that both the entropy and the energy scale identically with softness regardless of cutoff implies that the onset temperature is the same for rearrangements of all scales.

The fact that the energy experiences a shift that depends on the size of rearrangements is reminiscent of a similar shift in the energy of excitations, as a function of their size, that has previously been observed in glassy liquids[1]. In Fig. 5.6 we show that the dependence of the shift in the energy on the the cutoff is approximately logarithmic, in agreement with Ref.[1]. We therefore conclude that our choice of cutoff affects our results only through a shift of the the energy scale that can be accounted for by the fact that the rearrangement size depends linearly on the cutoff.

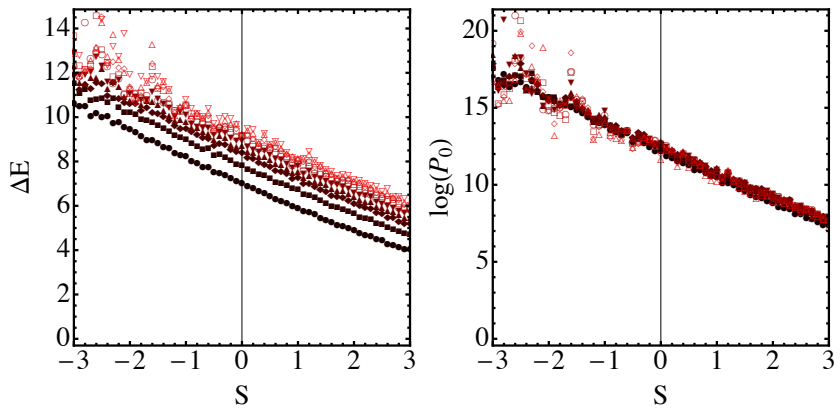


Figure 5.5: The dependence of the energy and entropy scales on the cutoff p_c. (a) The energy scale as extracted from the Arrhenius form for rearrangements with cutoffs p_c =0.05, 0.08, 0.11, 0.14, 0.17, 0.20, 0.23, 0.26, 0.29, 0.32 from black (lowest) to red (highest). (b) The entropy scale for the same cutoffs as in (a).

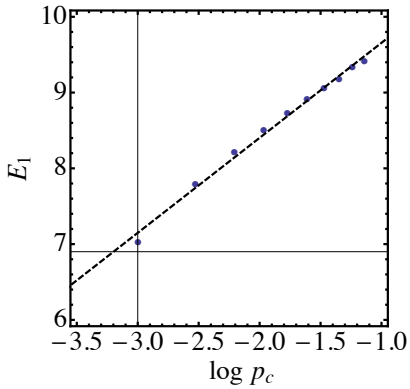


Figure 5.6: The shift in the energy scale as a function of the logarithm of the cutoff.

5.3.3 Computing softness.

We have made two improvements that greatly increased the prediction accuracy for rearrangements compared to the work on sheared systems in Chapter 4. First, we identified rearrangements more carefully, as detailed above. Second, we defined our training sets more carefully. Each training set contains 6000 particles that rearrange in the next time step, each labeled with $y_i = 1$, as well as 6000 particles that have

not rearranged for a time τ_α before the structure was calculated, each labeled with $y_i = 0$. These particles were chosen randomly from the set of all particles satisfying these conditions from MD simulations at a low temperature. Then, a training set of N particles can be written as $\{(\boldsymbol{F}_1, y_1), ..., (\boldsymbol{F}_N, y_N)\}$, where $\boldsymbol{F}_i = \left\{F_i^1, ..., F_i^M\right\}$ are the the M structure functions that describe the local neighborhood of particle i.

As in Chapter 4, we use two classes of functions to generate the set of M structure functions. The first class measures the density of particles a distance $r \pm \delta$ from a reference particle, i,

$$G_X(i; r, \delta) = \sum_{j \in X} e^{-\frac{1}{2\delta^2}(r - R_{ij})^2} \tag{5.2}$$

where R_{ij} is the distance between particles i and j and X denotes a species whose density we wish to probe. By varying r, δ, and X these functions investigate different aspects of the density of particles near the particle i. In this work we keep $\delta = 0.1\sigma_{AA}$ fixed. The second class of functions counts the number of small (or large) bond angles between triples of particles within a distance ξ of one another. It is given by,

$$\Psi_{XY}(i; \xi, \lambda, \zeta) = \sum_{j \in X} \sum_{k \in Y} e^{-(R_{ij}^2 + R_{jk}^2 + R_{ik}^2)/\xi^2} (1 + \lambda \cos\theta_{ijk})^\zeta \tag{5.3}$$

where θ_{ijk} is the angle between $\boldsymbol{R}_{ij}$ and $\boldsymbol{R}_{ik}$, $\lambda = \pm 1$ determines whether we consider small or large bond angles, ζ determines the angular resolution, and X and Y are species. By varying X, Y, ζ, λ, and ξ we probe different aspects of a particle's angular neighborhood.

We then use an SVM to find the hyperplane $\boldsymbol{w} \cdot \boldsymbol{F} - b = 0$ that separates the points with $y_i = 1$ from those with $y_i = 0$. This hyperplane is used on the rest of the data to reach the results reported. The SVM is trained, that is, the hyperplane is

constructed, on the binary variable y using the LIBSVM package[169]. It is not possible to find a hyperplane that perfectly separates the two different classes. We use a penalty parameter C and find the optimal hyperplane equation by minimizing

$$\frac{1}{2}\boldsymbol{w}^T \cdot \boldsymbol{w} + C\sum_{i=1}^{N}\chi_i, \tag{5.4}$$

with the constraint $y_i \cdot \left(\boldsymbol{w}^T \cdot \boldsymbol{F}_i + b\right) \geq 1 - \chi_i$ and $\chi_i \geq 0$ where χ_i are the slack variables. The C parameter was chosen through cross-validation. The hyperplane obtained from this training can be used to classify a new particle neighborhood, $\boldsymbol{F}_n$, as soft or hard. $\boldsymbol{F}_n$ is soft if $\boldsymbol{w} \cdot \boldsymbol{F}_n - b > 0$, and hard otherwise. The continuous variable softness is defined by $S_n = \boldsymbol{w} \cdot \boldsymbol{F}_n - b$. See supplemental information for other classification schemes explored.

5.4 Results

We can deduce the most important structural features contributing to softness either by training on fewer structure functions or by examining the projection of the hyperplane normal onto each orthogonal structure function axis. Both analyses yield a consistent picture (see Section 5.5): the most important features are the density of neighbors at the first peaks of the radial distribution functions $g_{AA}(r)$ and $g_{AB}(r)$; these two features alone give 77% prediction accuracy for rearrangements. Particles with more neighbors at the first peaks of $g(r)$ have a lower softness, and are thus more stable. These results are reminiscent of the cage picture, in which an increase of population in the first-neighbor shell suppresses rearrangements, or the free-volume picture, in which particles whose surroundings are closely-packed are more stable than those with more loosely-packed neighborhoods[184]. Overall, soft particles typically

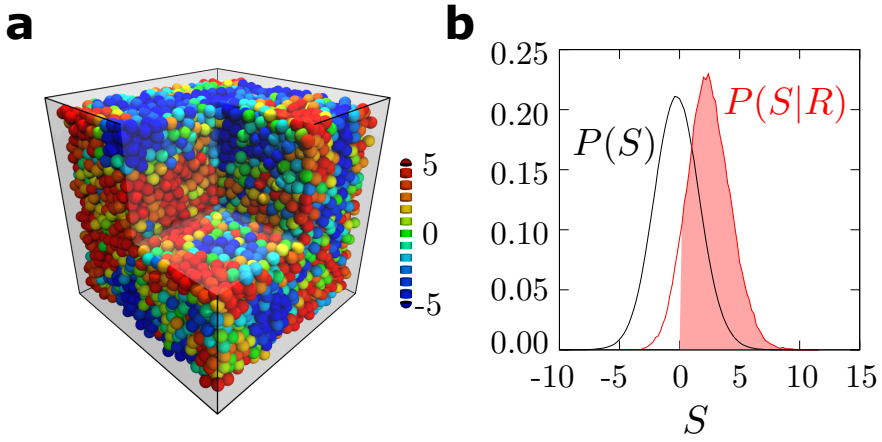


Figure 5.7: The characteristics of the softness field. a, A snapshot of the system at $T = 0.47$ and $\rho = 1.20$ with particles colored according to their softness from red (soft) to blue (hard). b, The distribution of softness of all particles in the system (black) and of those particles that are about rearrange (red). 90% of the particles that are about to rearrange have $S > 0$ (shaded region). None of the data included in this plot were in the training set.

have a structure that is more similar to a higher-temperature liquid, where there are more rearrangements, while hard particles whose structure appears closer to a lower-temperature liquid[185].

Fig. 1 (a) is a snapshot with particles colored according to their softness. Evidently, S has strong spatial correlations. Fig. 1 (b) shows the distribution of softness, $P(S)$, and the distribution of softness for particles just before they go through a rearrangement, $P(S|R)$. We see that 90% of the particles that undergo rearrangements have $S > 0$. We have also tested other sets of structure functions (see supplemental information) and found nearly identical accuracy. Softness is therefore a highly accurate predictor of rearrangements that is reasonably robust to the set of structure functions chosen.

We next show that the probability that particles rearrange is a function of their softness. This probability is calculated as the fraction of particles of a softness, S,

that are rearranging at a given time, $P_R(S)$. We plot $P_R(S)$ in Fig. 2 (a) in solid lines at temperatures ranging from $T = 0.47$ (blue) to $T = 0.58$ (red). At each T we see that $P_R(S)$ is a strong function of softness, increasing by several orders of magnitude, especially at the lower temperatures, in the range $S = -3$ to $S = 3$. A similar, albeit more modest, relationship was seen in[186]. When $P_R(S)$ is plotted as a function of $1/T$ for several values of softness, Fig. 2 (b), the probability that a particle of softness S will rearrange has Arrhenius behavior, $P_R(S) = P_0(S)\exp(-\Delta E(S)/T)$ where $P_0(S)$ and $\Delta E(S)$ depend on S. Confirming this observation, $P_R(S)/P_0(S)$ collapses over many orders of magnitude for all temperatures when plotted against $\Delta E(S)/T$, as shown in the inset of Fig. 2(b).

An Arrhenius form emerges when a kinetic process depends on a single energy scale $\Delta E(S)$. In Fig. 2 (c) we plot $\Delta E(S)$ and $\Sigma(S) \equiv \ln P_0(S)$ vs. S. Both terms depend nearly linearly on S: $\Delta E = e_0 - e_1 S$ and $\Sigma = \Sigma_0 - \Sigma_1 S$ where all four coefficients are positive and independent of T. Our results are consistent with the interpretation that at low temperatures, harder regions of the glassy liquid with higher energy barriers are frozen out while softer regions are not, leading to heterogeneous dynamics. These heterogeneities smooth out with increasing temperature, and vanish altogether once $P_R(S)$ no longer depends on softness. This occurs at the temperature T_0 where the softness dependence of Σ exactly cancels that of $\Delta E/T_0$ and so $T_0 = e_1/\Sigma_1$. This result can also be seen visually in Fig. 2 (b) where the different Arrhenius predictions for $P_R(S)$ all intersect at a single temperature, T_0, where the probability of rearrangement will be independent of softness. In Fig. 2 (d) we compare our prediction for T_0 to the onset temperature of glassy dynamics measured by Keys *et al.*[1], T_0^m, at different densities. The excellent agreement between the predicted T_0 and the measured values implies that the onset of glassy dynamics at $T = T_0$ coincides with the onset of

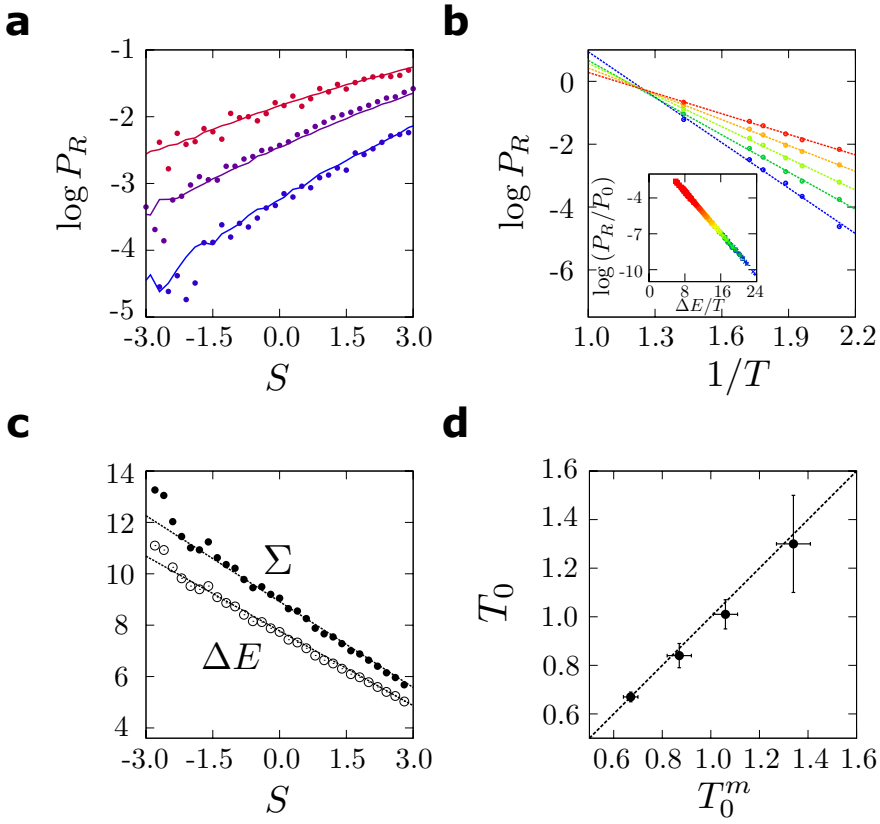


Figure 5.8: The relationship between softness and dynamics. a, The probability that particles rearrange as a function of their softness, $P_R(S)$, for temperatures T =0.47, 0.53, and 0.58 plotted in blue to red. Solid lines are measurements from molecular dynamics trajectories (solid lines). Dashed lines present the probability computed using the Arrhenius form for $P_R(S)$ (dashed lines). Points represent the probabilities calculated from the zero-time derivative of the overlap, $-dq(S,t)/dt$ at $T = 0.47$ and $T = 0.58$. b, $P_R(S)$ as a function of $1/T$ for 5 different softness values from $S \sim -3$ (blue) to $S \sim 3$ (red). The inset shows the collapse of these probabilities when P_R/P_0 is plotted against $\Delta E/T$. c, ΔE and Σ, where $P_R(S) = \exp(\Sigma - \Delta E/T)$, vs. softness S. d, predicted onset temperature T_0 vs. T_0^m, onset temperature measured by Keys, et al.[1], for densities $\rho = 1.15, 1.20, 1.25, 1.30$. The straight line corresponds to $T_0 = T_0^m$.

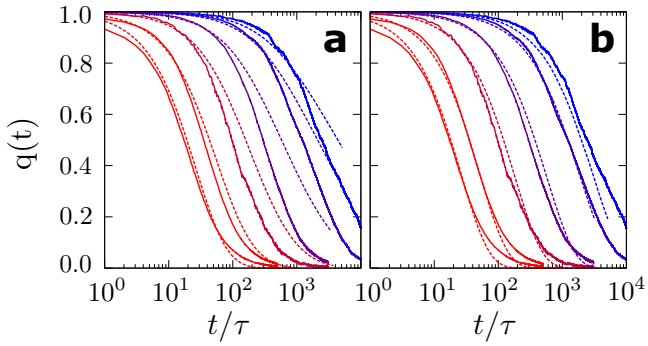


Figure 5.9: Overlap calculated from softness a, Solids lines are the measured overlap function, for temperatures $T = 0.45$, 0.47, 0.53, 0.58, 0.63, and 0.70, from blue to red, respectively. The dashed lines in the insets show predictions assuming each Arrhenius process is independent of one another. b, The solid lines in the insets are the same as in a. Dashed lines are predictions for the overlap function from $P_R(S)$ including changes in the softness field induced by spatial correlation between rearranging particles.

correlations between structure (softness) and dynamics.

We explore next the relationship between softness and the nonexponential decay of the overlap function

$$q(t) = \frac{1}{N} \sum_i \Theta(|r_i(t) - r_i(0)| - a)$$

where N is the number of particles in the system, r_i is the position of particle i, and Θ is the Heaviside function. We take $a = 0.5$[187]. In Fig. 3 (a)-(b) we plot the overlap function for different temperatures at $\rho = 1.20$. Our aim is to understand the form of the decay of $q(t)$ from the behavior of the rearrangement probability, $P_R(S)$. To begin, we define the contribution to the overlap from particles whose softness was initially S at $t = 0$, $q(S,t)$. The total overlap is $q(t) = \int dS q(S,t) P(S)$. Because $q(S,t)$ is the fraction of particles with initial softness S that have not rearranged after a time t, we expect $\frac{dq(S,t)}{dt}|_{t=0} = -c_a P_R(S)$ (see Section 5.6.1) where c_a is the fraction of rearrangements that displace particles by more than a. This is indeed the case, as is evident from the data in Fig. 2 (a), when $\frac{dq(S,t)}{dt}|_{t=0}$ (points) is overlaid with $P_R(S)$

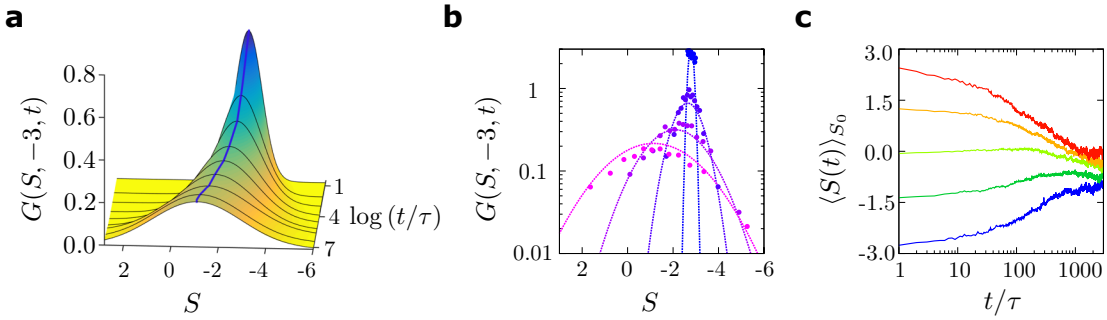


Figure 5.10: Time evolution of softness a, The stochastic evolution of softness in time as seen in through the evolution of the Gaussian approximation to the distribution of softness. b, The time evolution of the softness distribution for a collection of particles with initial softness $S_0 \sim -3$ from $t = 0$ (blue) to $t = 1000\tau$ (pink). Points are the measured histogram values, and the dashed lines are Gaussian approximations to the distribution. c, The time evolution of the average softness for particles that start from several softness values ranging from $S_0 \sim -3$ (blue) to $S_0 \sim 3$ (red).

(solid lines).

If we now assume that each particle rearranges with probability $P_R(S)$ as an independent Arrhenius process according to Fig. 2, then we can predict the decay of $q(S,t)$ using a simple discrete model: it can be written in terms of the probability that a particle of softness S does not rearrange for $t-1$ timesteps before finally rearranging at time t, $(1-P_R(S))^{t-1}P_R(S)$. The resulting prediction for $q(t)$ (dashed) is shown in Fig. 3 (a) for several different temperatures. While the prediction is not poor, its accuracy decreases at longer times, particularly at lower temperatures.

We now show that the discrepancy between our naive theory and the decay of $q(t)$ primarily results from a crucial neglected feature: even if a given particle does not rearrange, its local structural environment –and therefore its softness– can be altered if nearby particles rearrange. This physics is reminiscent of facilitation.

To take this facilitation into account, we calculate the "softness propagator", $G(S,S_0,t)$, that is, the distribution of softness at time t for particles that start with a softness S_0 at $t = 0$ and move less than a distance a after a time t (*i.e.* that do not rearrange

in a time t). Fig. 4 (a) shows a Gaussian approximation to $G(S, S_0 = -3, t)$. We see that $G(S, S_0, t)$ is sharply peaked around S_0 at small t but widens and shifts with increasing t reminiscent of directed diffusion. Fig. 4 (b) shows $G(S, S_0 = -3, t)$ at several different times, where points are measured probabilities and dashed lines are their Gaussian approximations. In Fig. 4 (c) we plot the mean softness evaluated as $\langle S(t) \rangle_{S_0} = \int dS S G(S, S_0, t)$ for several different values of S_0. For each S_0 the average softness of particles evolves towards the mean of the equilibrium softness distribution over a time period of approximately τ_α. The softness propagator is evaluated only for particles that have not rearranged, so Fig. 4 shows that rearrangements of nearby particles affect a particle's softness significantly.

Our first naive prediction based on the assumption that particles rearrange independently corresponds to $G(S, S_0, t) = \delta(S - S_0)$. We refine our theory by using the actual softness propagator in connecting the probability of rearranging, $P_R(S)$, with the overlap $q(S, t)$ (see Section 5.6.1). For ease of calculation, we approximate $G(S, S_0, t)$ as a Gaussian distribution in S and calculate its mean and variance as functions of S_0 and t from simulated data. The resulting prediction for the overlap is shown in Fig. 3 (b). The agreement with the actual $q(t)$ is excellent, suggesting that an understanding of the time evolution of the softness field, or equivalently of the softness propagator, would suffice to understand the non-exponential decay of the overlap function.

5.5 Feature selection and physical interpretation

The structure functions are inspired by the symmetry functions proposed by Behler and Parrinello[65] and angular structure functions are same. Our radial structure functions are different: we use functions that reduce to $r^{d-1} g(r)$ in the limit of $\delta \to 0$ in

Eq. 5.5, while they consider cumulative functions related to integrals of $g(r)$ from 0 to some distance R. We use 100 radial structure functions to describe a binary mixture of spheres, whereas as few as 16 radial symmetry functions have been used for such a system[65].

As discussed in Section 5.3.3, our radial structure functions are

$$G_X(i;r,\delta) = \sum_{j \in X} e^{-\frac{1}{2\delta^2}(r-R_{ij})^2} \tag{5.5}$$

where R_{ij} is the distance between particles i and j and X denotes the particle species whose density we wish to probe. We take r to be between 0 and 5.0 σ_{AA} in increments of $0.1\sigma_{AA}$, which gives a total of 50 radial structure functions per species. Our angular structure functions are

$$\Psi_{XY}(i;\xi,\lambda,\zeta) = \sum_{j \in X} \sum_{k \in Y} e^{-(R_{ij}^2+R_{jk}^2+R_{ik}^2)/\xi^2} (1+\lambda\cos\theta_{ijk})^\zeta \tag{5.6}$$

where θ_{ijk} is the angle between $\boldsymbol{R}_{ij}$ and $\boldsymbol{R}_{ik}$, $\lambda = \pm 1$ determines whether we consider small or large bond angles, ζ determines the angular resolution, and X and Y denote the particle species. Table 5.2 includes the angular parameters we have used in our work.

We now consider the physical meaning of softness. Using only the radial structure functions, a linear SVM with 88% cross-validation accuracy can be trained, as opposed to the 90% accuracy achieved by using both radial and angular structure functions. Despite the 2% loss in accuracy, we focus the remaining analysis in this section on only the radial functions, since their physical interpretation is more straightforward (i.e. $G_X(i;r,\delta)$ counts the number of particles of species X within δ of r.) We

Structure function parameters for Ψ_{YZ}			
	ξ (σ_{AA})	ζ	λ
1	14.633	1	-1
2	14.633	1	1
3	14.638	2	-1
4	14.638	2	1
5	2.554	1	-1
6	2.554	1	1
7	2.554	2	-1
8	2.554	2	1
9	1.648	1	1
10	1.648	2	1
11	1.204	1	1
12	1.204	2	1
13	1.204	4	1
14	1.204	16	1
15	0.933	1	1
16	0.933	2	1
17	0.933	4	1
18	0.933	16	1
19	0.695	1	1
20	0.695	2	1
21	0.695	4	1
22	0.695	16	1

Table 5.2: Angular structure function parameters.

begin by computing softness using various subsets of the radial functions to determine which are most important. We will support these results by considering the weights identified by the SVM trained on all of the radial functions. Overall, we show that fluctuations in local density at the peaks of $g(r)$ have the most influence on softness.

5.5.1 Using a single radial structure function

We train a linear SVM on each of the 100 radial functions separately for a training set of type A particles. In Fig. 5.11, we plot the accuracy of these classifiers, along with the radial distribution function. For SVMs trained on type B structure functions (corresponding to $g_{AB}(r)$), the highest accuracy is at the first peak of the radial

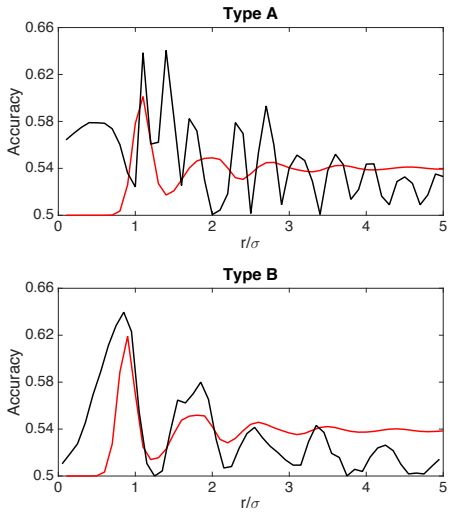


Figure 5.11: Black curves represent the classification accuracy of the SVM trained on a radial structure function at r. Red curves are the radial distribution function of neighbors of the given type.

distribution function. The accuracy of the classifier roughly follows the normalized density of type A neighbors at the distance the structure function is calculated. For classifiers trained on type A structure functions (corresponding to $g_{AA}(r)$, the results are harder to interpret. The highest accuracy is achieved at the first peak of the radial distribution function, as was the case for type A. However, a similarly sized peak is seen at the first trough, and the accuracies do not seem to follow the normalized density of the neighbors.

The radial structure functions around the first peak of the radial distribution function seem to be most important, based on the results of single structure function trainings. No single structure function can achieve a classification accuracy higher than 64%.

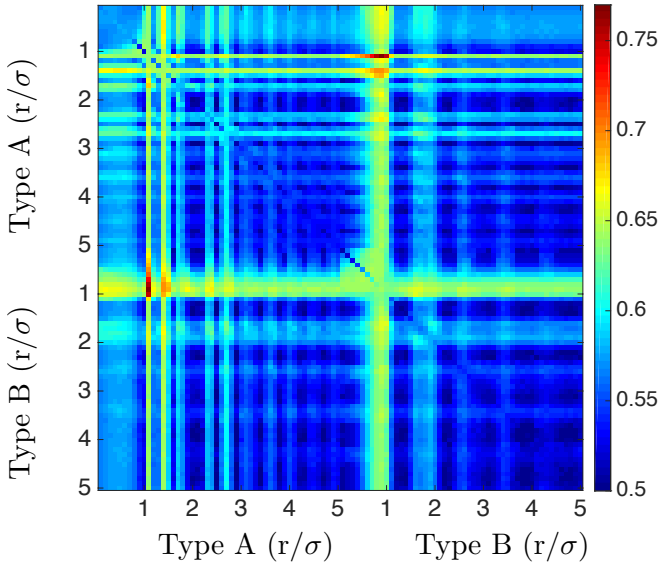


Figure 5.12: The accuracy of an SVM trained on a pair of radial structure functions. The axes denote type / distance of the structure functions used and the color denotes the resulting cross-validation accuracy.

5.5.2 Using two radial structure functions

Next, we train a linear SVM on each pair of radial structure functions using a training set of type A particles. In Fig. 5.12, the accuracy for each pair is presented. The highest accuracy (77%) is achieved by using the type B function at $0.9\sigma_{AA}$ and the type A function at $1.1\sigma_{AA}$, both of which correspond to the first peaks of their respective radial distribution functions. Thus, an accuracy of 77% is obtained if 2 quantities are considered, the height of the first peak of the AB and of the AA radial density functions.

5.5.3 Using three radial structure functions

We trained a linear SVM on all triplets of radial structure functions and recorded their accuracy. The triplet with the best accuracy (78%) has the two structure func-

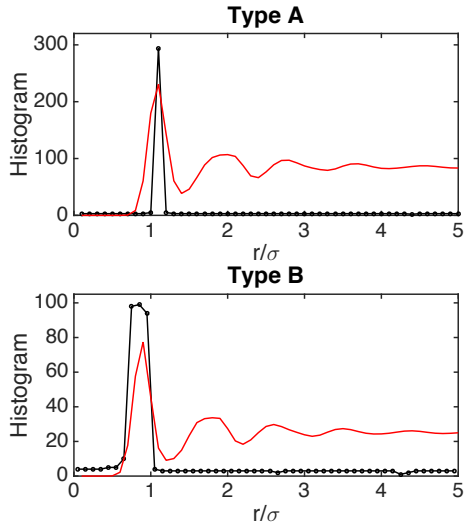


Figure 5.13: Black curves denote the number of high accuracy triplets a radial structure function was in. Red curves denote the radial distribution function in arbitrary units.

tions from the best pair (from the previous subsection), corresponding to the first peaks of the AB and AA radial distributions, in addition to the type B function at $1.9\sigma_{AA}$, which is at the second peak of the AB radial distribution function.

Out of 161,700 triplets of radial structure functions, only 296 of them can be used to achieve an accuracy above 75%. 294 of these 296 triplets include the first peak of the AA radial distribution (type A function at $1.1\sigma_{AA}$), and 291 of the 296 triplets have one of the type B functions at 0.8, 0.9, or $1.0\sigma_{AA}$ (the first peak of the AB radial distribution). These results are summarized in Fig. 5.13.

Next, we can analyze the weights these structure functions have when they are in a good triplet. Fig. 15 shows the median weight of the structure functions when they are in a triplet of structure functions with better than 75% accuracy. We see that the largest magnitude weights correspond to the first peaks of the radial distribution functions, and they are negative. This means that having more neighbors at the first

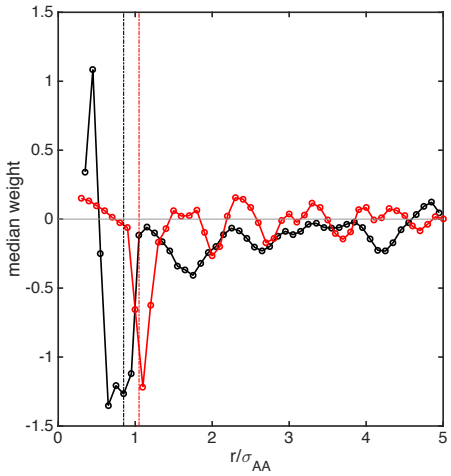


Figure 5.14: Red (black) curve shows the median weight assigned to the type AB (BB) structure function when they are in a high accuracy triplet.The dashed lines denote the first peak location.

peak makes a particle harder.

5.5.4 Using more than three radial structure functions

It is prohibitively time-consuming to calculate the accuracy of all combinations with more than five structure functions. For combinations of up to $N_s = 15$ structure functions, we used evolutionary algorithms to find the optimal set of structure functions. At least 2 million combinations were evaluated for the optimization for each value of N_s. The results are summarized in Fig. 5.15. There is a large accuracy gain from using two radial structure functions rather than one. When $N_s = 15$ structure functions are used, the accuracy is 86%, very close to the accuracy of 88% that is obtained when all 100 radial structure functions are used.

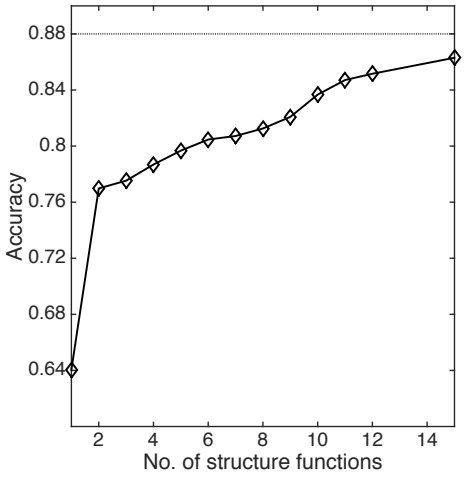


Figure 5.15: Best accuracy achieved for a given number of radial structure functions. The dashed lines represent the accuracy achieved by using all 100 radial structure functions.

5.5.5 Weight analysis from training with all radial structure functions

An alternative way to understand the importance of the various features is to look at the hyperplane normal after training. Recall that the softness is given by,

$$S_i = \sum_{\alpha} w_\alpha \tilde{G}(i; r_\alpha, \delta) \tag{5.7}$$

where w_α is the hyperplane normal for the radial structure function centered at r_α and $\tilde{G} = (G - \langle G \rangle)/\sqrt{\langle \delta G^2 \rangle}$ is the normalized structure function with zero mean and unit variance. Thus, $\tilde{G}(i; r, \delta)$ is the normalized fluctuations in density from the average density at a distance r. It follows that density fluctuations at a radius with a larger weight will have a greater impact on softness (and are therefore more important in defining softness.) In Fig. 5.16 we plot the normal components as a function

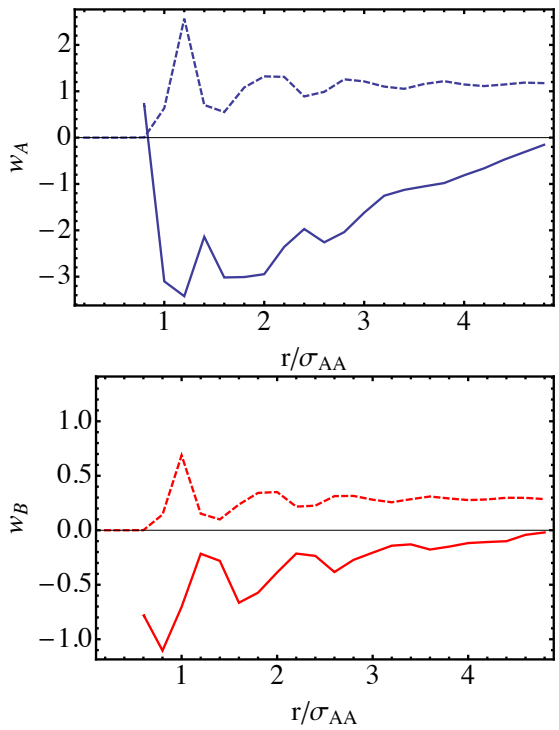


Figure 5.16: Solid lines are the weights associated with the radial structure functions. Dashed lines are the radial distribution functions.

of r/σ_{AA} overlaid with $g(r)$. We see that the most important symmetry functions reside at the successive peaks of $g(r)$ with the troughs contributing more modestly to softness.

5.5.6 Physical interpretation of the weights

In the previous sections we have analyzed which weights are most important in determining softness. We now discuss the physical implications of these results. First we notice in Fig. 12,13,14,15, and 17 that the most important features that contribute to the value of softness for an A-particle are consistently the density of neighbors at the first peak of the radial distribution functions, $g_{AA}(r)$ and $g_{AB}(r)$. The weights for

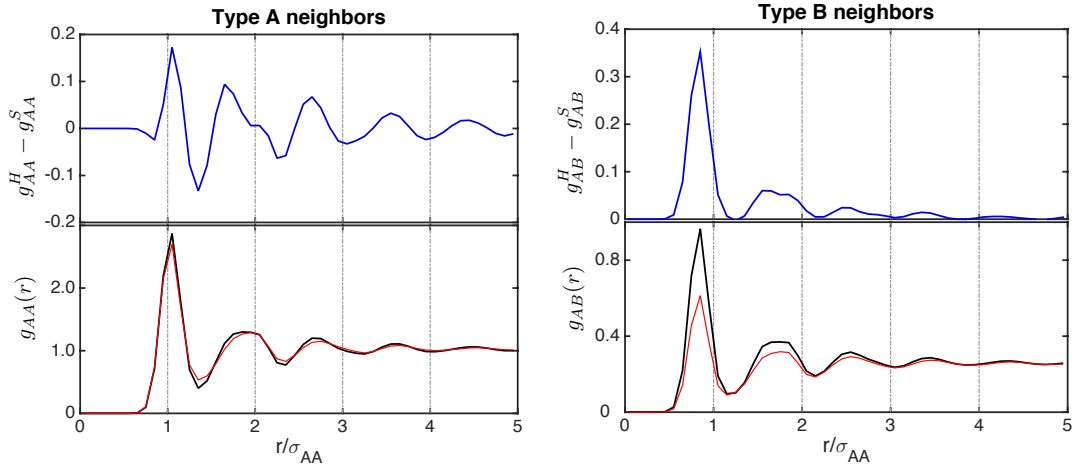


Figure 5.17: Left (right) panels refer to the radial distribution function for type A (B) neighbors. Bottom panels show the radial distribution functions for hard (black) and soft(red) particles. Top panels show the difference between these two radial distribution functions.

the first peak are always negative, implying that particles whose local environments are more tightly packed are harder. This is in agreement with our intuition from free-volume theory or the cage description of glasses, in which particles with more tightly-packed neighborhood should exhibit fewer hops[184].

To see the importance of the first peaks of the radial distribution functions directly, we consider the functions separately for hard and soft central particles respectively. For neighbors of species A and B we plot, in Fig. 18, the distribution separately for hard and soft particles, denoted g^H_{AX} and g^S_{AX} as well as the difference between these two functions. We see that for both types of neighbors, the difference between g^H_{AX} and g^S_{AX} is largest at the first peak of $g_{AX}(r)$. This description is corroborated by looking at Fig. 12. Here we see that again the first peaks of $g_{AX}(r)$ are consistently the most important in determining softness.

Considering Fig. 18 we see that hard and soft particles are distinguished qualitatively by local environments that come from lower and higher temperature liquids re-

spectively. This is in agreement with our earlier findings in two-dimensions(see Chapter 4) and provides an appealing and intuitive explanation for softness.

5.5.7 Alternative structure functions

As noted above, our original choice of structure functions was inspired by the symmetry functions proposed by Behler and Parrinello[65], and we viewed these structure functions as capturing generic two- and three-point correlations in the local environment of a given particle. One could ask whether this is the correct view, or if instead the Behler-Parrinello functions give a privileged or optimal representation of the local structure. To gain some insight, we have constructed alternate choices for the both the radial and angular structure functions.

An alternate representation of the radial environment is a local version of the cumulative radial distribution. We define the associated structure functions for particle i to be

$$G^Y_{cdf}(i;\mu) = \sum_j 1, \qquad R_{ij} < \mu. \tag{5.8}$$

That is, $G^Y_{cdf}(i;\mu)$ simply counts the number of particles of species Y that are within μ of particle i. As before we vary Y over both species of particles, and we choose the μ to be evenly spaced distances between 0 and $3.5\sigma_{AA}$ with a spacing of $0.1\sigma_{AA}$.

For the angular functions we switched to structure functions based on the spherical harmonics, $Y_{lm}(\vec{r})$. Following Steinhardt et al.[188] we define the following angular functions:

$$Q_l(i; r_{min}, r_{max}) = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |\langle Q_{lm}(\vec{r})\rangle|^2 \right)^{1/2}, \tag{5.9}$$

where

$$\langle Q_{lm}(\vec{r})\rangle = \frac{1}{N_j}\sum_j Y_{lm}(\vec{R}_{ij}), \qquad r_{min} < R_{ij} < r_{max}. \tag{5.10}$$

$\langle Q_{lm}(\vec{r})\rangle$ is thus the average value of the $Y_{lm}(\vec{r})$ for every particle j in a shell near particle i, and $Q_l(i; r_{min}, r_{max})$ forms a rotationally invariant combination of the Q_{lm}'s. For our structure functions we chose r_{min} and r_{max} to be shells of width $0.5\sigma_{AA}$ where the inner radius started at 1.0, 1.5, 2.0, 2.5, and 3.0 (all in units of σ_{AA}). We chose l parameters in the set $l \in \{2, 4, 6, 8, 10, 12, 14\}$.

To test these alternate structure functions, we chose a 10000-particle training set at $\rho = 1.15$, $T = 0.37$ and a 6000-particle training set at $\rho = 1.2$, $T = 0.47$. We applied each pairwise combination for the radial ($g(r)$-like or CDF-like) and angular (Behler-Parrinello or Q_l) structure functions and measured the cross-validation accuracy obtained by subdividing the training sets into six smaller sets. In all cases we found cross-validation accuracies that were within 2% of those found in Section 5.4. Thus, our original set of structure functions is complete enough to provide high accuracy. It is not optimal, because alternate sets of structure functions work equally well.

Choosing a larger spacing for the G^Y_{cdf} modestly degrades the cross-validation accuracy, but we found no gains by going to a finer resolution. We also investigated the effects of separately defining the Q_l functions for A-type and B-type particles, but found no substantial improvement in our accuracy. For the angular functions we also tried including the third-order invariants

$$\hat{W}_l = W_l \left(\sum_{m=-l}^{l} |\langle Q_{lm}(\vec{r})\rangle|^2 \right)^{-3/2}, \tag{5.11}$$

where

$$W_l = \sum_{m_1,m_2,m_3} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} (\langle Q_{lm_1}(\vec{r})\rangle \langle Q_{lm_2}(\vec{r})\rangle \langle Q_{lm_3}(\vec{r})\rangle) \tag{5.12}$$

and the coefficients

$$\begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{5.13}$$

are the Wigner $3j$ symbols. However, we again found that for our training sets including the W_l's (over the same range of l as in the Q_l's) did not measurably change our cross-validation accuracy.

In addition, we tried switching or augmenting our radial structure functions with the original radial symmetry functions[65]. Furthermore, we tried augmenting our features with ordered distances of the nearest 160 neighbors of a particle. None of these alternate features measurably improved the cross-validation accuracy, which leads us to believe that our set of structure functions provides a reasonably complete description of the local structural environment.

5.6 Discussion

5.6.1 The relationship between P_R and the overlap

We construct a model to relate the probability of rearrangement $P_R(S)$ to the softness-dependent overlap function

$$q(S,t) = \frac{1}{N_S} \sum_i \Theta(|r_i(t) - r_i(0)| - a)\delta(S_i(0) - S) \tag{5.14}$$

where N_S is the number of particles of softness S at $t = 0$. As noted in the above, if all rearrangements are irreversible, this overlap will be identically equal to the fraction of particles that have not rearranged after a time t. We consider a discrete model where time advances in units of $\tau_R(T)$ so that at each timestep particles may either rearrange or not rearrange and each rearrangement is completed in approximately one timestep. Since $P_R(S)$ potentially includes rearrangements that are reversible or result in particle displacements whose magnitudes are less than a, we use a rescaling $P_R(S) \to c(T)P_R(S)$. Here $c(T)$ is a function that represents the fraction of rearrangements that contribute productively to relaxation at the lengthscale a and is given by approximately $c(T) = c_a f_{irrev}(T)$ where c_a is the fraction of rearrangements that result in displacements greater in magnitude than a and $f_{irrev}(T)$ is the fraction of irreversible rearrangements discussed above.

Using the softness propagator $G(S, S_0, t)$, we write the fraction of particles that start with a softness S_0 at $t = 0$ and rearrange after a time t as

$$f_t = \int dS G(S, S_0, t) P_R(S). \tag{5.15}$$

It follows that the fraction of particles that have not rearranged up to a time $t - 1$ before finally rearranging after a time t will be,

$$P(t|S_0) = \prod_{t'=0}^{t-1}(1 - f_{t'})f_t = \prod_{t'=0}^{t-1}\left[\int dS G(S, S_0, t')(1 - P_R(S))\right]\int dS G(S, S_0, t) P_R(S) \tag{5.16}$$

The fraction of particles that have not rearranged after a time t – and hence the value

of the overlap function – will be given by,

$$q(S_0,t) = 1 - \sum_{t'=0}^{t} \prod_{t''=0}^{t'-1} \left[\int dS G(S,S_0,t'')(1-P_R(S)) \right] \int dS G(S,S_0,t') P_R(S). \quad (5.17)$$

This gives a softness dependent prediction for the overlap that can be computed analytically once $P_R(S)$ and $G(S,S_0,t)$ are known. The overall overlap is related to $q(S,t)$ by

$$q(t) = \int dS q(S,t) P(S). \quad (5.18)$$

In Fig. 5.18 (a) we see the overlap as a function of time for different softnesses at two representative temperatures. In both cases we see good agreement between the predicted and measured values of $q(S,t)$. Similarly in Fig. 5.18 (b) we see the average overlap for all temperatures considered and find similarly good agreement.

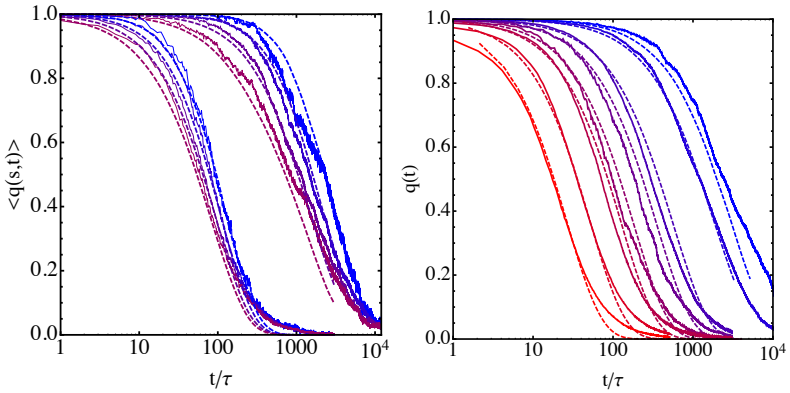


Figure 5.18: The non-exponential decay of overlap. (a) the softness-dependent overlap $q(S,t)$ for two representative temperatures $T = 0.47$ (long time) and $T = 0.58$ (short time) at four softnesses from -4 (blue) to 4 (red). (b) the average overlap at all temperatures from $T = 0.45$ (blue) to $T = 0.70$ (red).

When computing $q(S,t)$ we approximate $G(S,S_0,t)$ by a Gaussian because it allows us to convolve the probability of rearrangement, $P_R(S)$, with the propagator analyt-

ically. This is important given that it can be difficult to get enough data to numerically integrate $G(S, S_0, t)$ against $P_R(S)$ since $G(S, S_0, t)$ is defined only for those particles that have not yet rearranged.

However, instead of computing the softness dependent overlap, $q(S, t)$, and then integrating over $P(S)$ to find the total overlap $q(t)$, one can alternatively compute the total overlap directly:

$$q(t) = 1 - \sum_{t'=0}^{t} \left[\prod_{t''=0}^{t'-1} \int dS (1 - P_R(S)) P(S, t'') \right] \int dS P_R(S) P(S, t')$$

where we define $P(S, t) = \int dS_0 G(S, S_0, t) P(S_0)$. Here $P(S, t)$ is full softness distribution for particles that have not rearranged after a time t. While this formulation obscures the dynamics of the softness field, we have sufficient data to reconstruct the full distribution of $P(S, t)$ from simulation, thus avoiding the Gaussian approximation. The results of this direct calculation agree well with those calculated using the Gaussian approximation, so we conclude that the Gaussian approximation does not skew dynamics of the average overlap substantially. However, individual overlaps, $q(S, t)$, might be under- or over-estimated by the Gaussian approximation if the true propagator has a substantial amount of skew.

5.6.2 Other classification schemes

The results in this chapter were obtained using a linear SVM kernel. However, we also trained SVM classifiers using the radial basis function (RBF) kernel. There are two parameters associated with this kernel, γ and C. We search over the optimal parameters over a fine grid, using a training set of 10,000 points. All of the structure functions were used as features (100 radial and 66 angular). The features are stan-

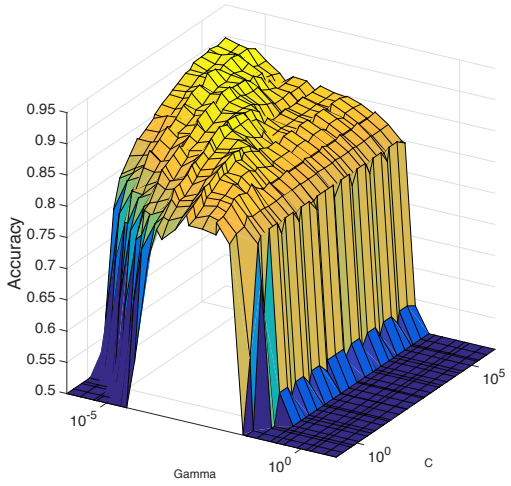


Figure 5.19: Cross-validation accuracy as a function of γ and C.

dardized before training.

The cross-validation accuracy using the optimal parameters was 90.4% (Fig. 5.19), which is equal to the cross-validation accuracy achieved using the linear kernel. For this reason, we use the linear kernel for its simplicity.

In addition, we trained neural network classifiers for softness. Using a training set of 10,000 points, we trained networks with one or two hidden layers and one to fifty hidden nodes in each layer. All of the networks had a validation accuracy below 90%, below the accuracy obtained with the simple linear SVM kernel.

5.6.3 Nonlinear softness dependence of ΔE and $\log P_0$

In Section 5.4 we show that the energy scale and the entropy depend nearly linearly on softness. At large values of S, however, both the energy and the entropy deviate slightly from the linear prediction. This effect is shown in Fig. 5.20. Intuitively, this must be the case as there must be a lowest energy barrier in the system and this

energy barrier must be positive by construction. Mathematically, this must also be the case since the probability of a rearrangement can never exceed one. This effect becomes important, in particular, when we compute the average probability of rearrangement by integrating the product of $P_R(S)$ and a Gaussian distribution. In this case, if the deviation from linear scaling is neglected, the integral can become dominated by the unphysical regime where the distribution is small but $P_R(S) \gg 1$. To

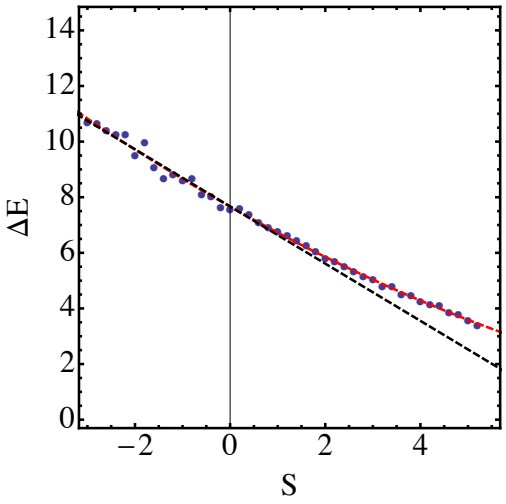


Figure 5.20: The deviation of the energy scale $\Delta E(S)$ from linear behavior in S at $T = 0.47$ and $\rho = 1.20$.

correct this problem we do not make any assumptions about the functional form of $\Delta E(S)$ and $\log P_0(S)$ at large S. We instead fit the data including a quadratic term, $E(S) = E_0 + E_1 S + E_2 S^2$ to ensure convergence of integrals. Typically E_2 is about an order of magnitude smaller than E_1.

5.6.4 The fraction of irreversible rearrangements

Now we examine how the fraction of irreversible rearrangements changes with temperature. To compute the fraction of rearrangements of a scale $a = 0.5$ that are irre-

versible we begin by computing a running average of particle displacements from their $t = 0$ positions. We choose the averaging window such that it is large compared to $\tau_R(T)$ but its size has a very weak impact on the results. We then count what fraction of particles' average displacements exceed a and then return to within a distance $b = 0.2$ of their starting position. Again, the results depend very weakly on the choice of b. These are the reversible rearrangements.

In Fig. 5.21 (a) we show the fraction of rearrangements that are irreversible, f_{irrev} as a function of softness at different temperatures. We see that at all temperatures f_{irrev} is approximately independent of softness. Fig. 5.21 (b) shows that as temperature is lowered towards the glass, f_{irrev} decreases markedly. These results are consistent with the observations of Vollmayr-Lee[189] who saw a similar decrease in f_{irrev} upon cooling.

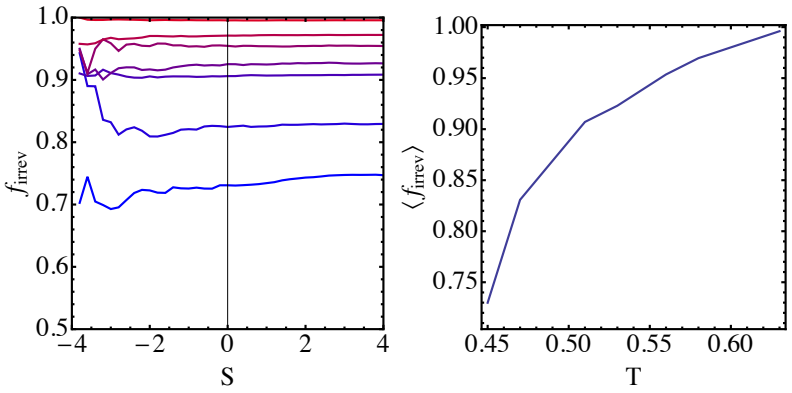


Figure 5.21: The fraction of rearrangements that are irreversible, f_{irrev} (a) as a function of softness for temperatures from $T = 0.47$ (blue) to $T = 0.63$ (red); (b) as a function of temperature.

5.7 Conclusion

Our results show that there is structure hidden in the disorder of glassy liquids. This structure can be quantified by softness, which controls glassy dynamics at tempera-

tures below T_0. According to our analysis, simple Arrhenius relaxation for each softness, coupled with the time evolution of softness, leads to the observed slow, non-exponential relaxation dynamics of glassy liquids below T_0. Thus, our results suggest that the challenge of understanding glass transition dynamics can be reframed as the challenge of understanding the evolution of softness.

6

Structural Properties of Defects in Glassy Liquids

At zero temperature a disordered solid corresponds to a local minimum in the energy landscape. As the temperature is raised or the system is driven with a mechanical load, the system explores different minima via dynamical events in which particles rearrange their relative positions. We have shown in the previous chapters that the dynamics of particle rearrangements are strongly correlated with a structural quantity associated with each particle–"softness"–which we can identify using supervised machine learning. Particles of a given softness have a well-defined energy scale that

governs local rearrangements; because of this property, softness greatly simplifies our understanding of glassy dynamics. Here we investigate the correlation of softness with other commonly used structural quantities, such as coordination number and local potential energy. We show that although softness strongly correlates with these properties, its predictive power for rearrangement dynamics is much higher. We introduce a useful metric for quantifying the quality of structural quantities as predictors of dynamics. We hope that in the future, authors introducing new structural measures of dynamics will compare their proposals quantitatively to softness using this metric. We also show how softness correlations give insight into rearrangements. Finally, we explore the physical meaning of softness using unsupervised dimensionality reduction and reduced curve-fitting models, and show that softness can be recast in a form that is amenable to analytical treatment.

6.1 Introduction

In a crystal, rearrangements of the constituent particles due to thermal fluctuations or an applied mechanical load occur almost exclusively at structural defects such as dislocations. In a high-temperature liquid, by contrast, there are no special structural defects and any of the constituent particles can rearrange at any time. It has therefore not been obvious whether rearrangements in glasses or in liquids near the glass transition occur at structural defects of some kind. The clear gain in simplification provided by a defect description has driven successful phenomenological approaches[148,190] that have posited the existence of defects correlated with rearrangements. It is only in the last several years, however, that such defects have been identified unambiguously[43,159,191,192]. In the previous two chapters, we have shown that machine learning (ML) methods provide an efficient means for identifying structural

defects associated with rearrangements with remarkably high accuracy. Using ML, we identify a scalar quantity, "softness," for each particle that is a function of its local structural environment and is strongly correlated with the probability that the particle will exhibit a rearrangement. The concept of softness has already proven very useful in simplifying our understanding of dynamics of glassy liquids. It is therefore important to understand the physical interpretation of softness in greater depth.

In this chapter we explore the properties of softness in a number of ways. We begin by considering the relationship between softness and other commonly used measures of local structure including local potential energy, coordination number, and topological clustering. We find that while these quantities correlate well with softness, they are poor predictors of dynamics, either alone or in combination.

Next we turn to softness more directly and investigate spatial correlations of softness and show that softness is correlated only on the scale of the particle size and that fluctuations in softness are strongly correlated with particle motion during rearrangements.

We then manipulate the definition of softness mathematically to shed light on its physical interpretation. Here we show that softness can be written as a sum over pairwise contributions. We also identify several approximations to softness that involve far fewer variables and are easier to work with from a theoretical standpoint, and yet are nearly as accurate as softness in predicting rearrangements.

Finally, we use a class of ML techniques to perform so-called "unsupervised dimensionality reduction". Unlike the methods that we employ to construct softness, these techniques simply try to cluster structurally similar points together in a two-dimensional space. Despite this relatively general ansatz, we find that this clustering correlates strongly with softness and sheds light on how defects are manifested in dis-

ordered solids.

6.2 Background

The work presented here is based on studies of two systems: a two-dimensional sheared Kob-Andersen Lennard-Jones glass below the glass transition temperature and a three-dimensional quiescent Kob-Andersen Lennard-Jones supercooled liquid. Detailed treatments of these systems may be found in Refs.[1,182,193]. Throughout the chapter, distances are measured in units of the large particle diameter, σ_{AA}, energies in units of ϵ_{AA} and mass in terms of the common mass of the two particles, m. The two-dimensional system is at a temperature of $kT/\epsilon = 0.1$ and shear strain rate $\dot{\gamma} = 10^{-4}$ while the three-dimensional system is at a temperature of $kT/\epsilon = 0.47$. Both systems have 10,000 particles and are at number density $\rho = 1.2$. In both cases the systems were studied using molecular dynamics simulations and a Nosé-Hoover thermostat was used to control the temperature. For the $d = 3$ system, the configuration of the system was recorded and the system was quenched to its nearest inherent structure using the FIRE algorithm[194] every τ, where $\tau = \sqrt{m\sigma_{AA}^2/\epsilon_{AA}}$ is the time in simulation units. The configuration was saved and quenched every 2τ for the d=2 system.

Particle dynamics consist of small "caged" fluctuations punctuated by larger cooperative rearrangements. To identify rearrangements in the three-dimensional quiescent system we use a "hop" indicator function, p_{hop}, that was introduced by Candelier *et al.*[155,195]. To define p_{hop} at a time t we first specify a duration, Δt, as well as two intervals $A = [t - \Delta t/2, t]$ and $B = [t, t + \Delta t/2]$. We may then write,

$$p_{\text{hop}}(t) = \sqrt{\langle(\boldsymbol{r}_i - \langle\boldsymbol{r}_i\rangle_B)^2\rangle_A \langle(\boldsymbol{r}_i - \langle\boldsymbol{r}_i\rangle_A)^2\rangle_B} \tag{6.1}$$

where $\langle\rangle_A$ and $\langle\rangle_B$ denote averages over the A and B intervals respectively. During caged motion $\langle \boldsymbol{r}_i \rangle_A = \langle \boldsymbol{r}_i \rangle_B$ and p_{hop} will be commensurate with the variance in the particle position due to noise in the inherent structure calculation. During rearrangements the mean position will change markedly and p_{hop} will attain significantly larger values. We take rearrangements to be those events with $p_{\text{hop}} > p_c = 0.2$. This choice is discussed at length in the previous chapter.

In the two-dimensional sheared systems a large affine background makes p_{hop} an unreliable indicator of rearrangements. In this system we therefore use the total non-affine displacement of a region around a particle i. This measure, first introduced in Ref.[149], is defined by

$$D^2_{\text{min}}(i) = \sum_{j \text{ s.t. } R_{ij}<R_c} \left[\boldsymbol{R}_{ij}(t+\Delta t) - \boldsymbol{\Lambda}_{\text{min}} \boldsymbol{R}_{ij}(t)\right]^2 \tag{6.2}$$

where Λ is the best fit affine transformation. Here we take $R_c = 2.5\sigma$ and $\Delta t = 5\tau$; we note, however, that our results are qualitatively insensitive to these choices. It has been shown that D^2_{min} is a reliable indicator of plastic rearrangements in disordered systems under applied strain. Again we see a clear separation of scales between the non-affine displacement of caged motion and rearrangements.

In Chapter 5 we showed that the softness S can be directly related to the probability that particles rearrange so that,

$$P_R(S) = \exp\left[\Sigma(S) - \frac{\Delta E(S)}{kT}\right] \tag{6.3}$$

where $\Sigma(S) = \Sigma_0 - \Sigma_1 S$ and $\Delta E(S) = e_0 - e_1 S$. Thus, there is a well-defined energy barrier associated with rearrangements of particles of a given softness; in other words,

softness is a measure of the energy barrier that governs rearrangements of particle i. From $\Sigma(S)$ and $\Delta E(S)$ one can calculate the onset temperature, T_0, below which a liquid develops structural defects associated with rearrangements. This is also the temperature below which the liquid develops glassy signatures such as stretched exponential relaxation, a non-Arrhenius dependence of relaxation time on temperature and kinetic heterogeneities.

6.3 Softness and related quantities

In Chapter 5 we showed that for the $d = 3$ Kob-Andersen liquid we could limit our structure functions to those of the form[65],

$$G_X(i; r, \sigma) = \frac{1}{\sqrt{2\pi}} \sum_{j \in X} e^{-\frac{1}{2\sigma^2}(R_{ij}-r)^2} \tag{6.4}$$

with almost no loss in predictive accuracy. Here, R_{ij} is the distance between particles i and j. Functions of this form count the number of particles a distance $r \pm \sigma$ away from a target particle i. Thus, what matters to softness in this system is the value of the pair correlation function $g(r)$ at different values of r, especially at the peaks and valleys (Chapter 5.5).

Here we look at the correlation between softness and other widely-used characterizations of the structure of the local environment of a particle. It turns out that softness correlates with many physical quantities, such as the local potential energy, number of particles in the first neighbor shell, and geometrical measures of the neighborhood. However, we find that none of these quantities is nearly as predictive of rearrangements as softness, either on their own or in combination with each other.

We created a data set of 160,000 particles, randomly chosen from a molecular dy-

namics simulation trajectory of the glassy $d = 3$ Kob-Andersen liquid. For each particle in this set, we calculated the softness, the number of particles in the first neighbor shell and the local potential energy. The local potential energy for a particle is defined as the sum of the Lennard-Jones interactions of the particle with its neighbors, calculated using the same cutoff, $2.5\sigma_{AA}$, that was used for the calculation of structure functions. As seen in Fig. 6.1(a), softness is anti-correlated with the number of neighbors in the first shell. This is expected, since we found in Chapters 4 and 5 that soft particles have more liquid-like pair correlation functions. Particles with fewer than 13 neighbors are soft on average, while particles with more than 13 neighbors are hard on average. However, only 60% of rearranging particles have fewer than 13 neighbors. Thus, coordination in the first shell alone is not an effective predictor of rearrangements.

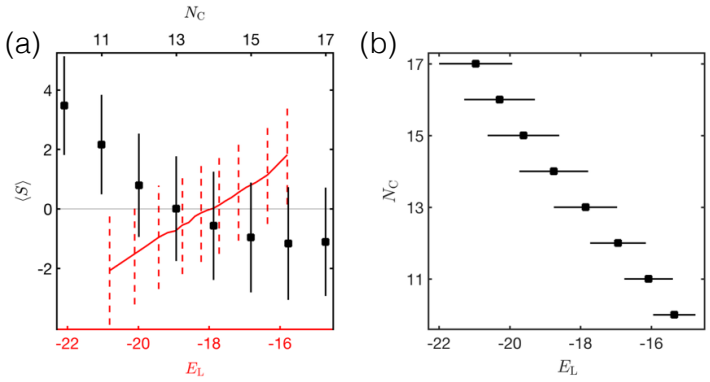


Figure 6.1: (a) Average softness vs. coordination number is shown with black squares, with the corresponding x axis at the top. Average softness as a function of local energy is calculated with a fixed bin size of 8000 points, drawn with a solid red line, with the x axis at the bottom. The red and black error bars represent the standard deviation of softness for corresponding points as a function of local energy and coordination number, respectively. (b) Local energy vs. coordination number. The error bars represent the standard deviation of local energy of particles with corresponding coordination number.

Similarly, Fig. 6.1(a) also shows the correlation of softness with the local poten-

tial energy. Evidently softness is linearly correlated with local energy. This was also observed by Royall et al., who found that energetically locally favoured structures are longer lived[181]. We see that this positive correlation between local energy and softness holds for all local potential energies and that particles with local energies $E_L > -18$ are soft on average. However, only 65% of rearrangements occur at particles with $E_L > -18$. The local potential energy is not nearly as predictive of rearrangements as softness.

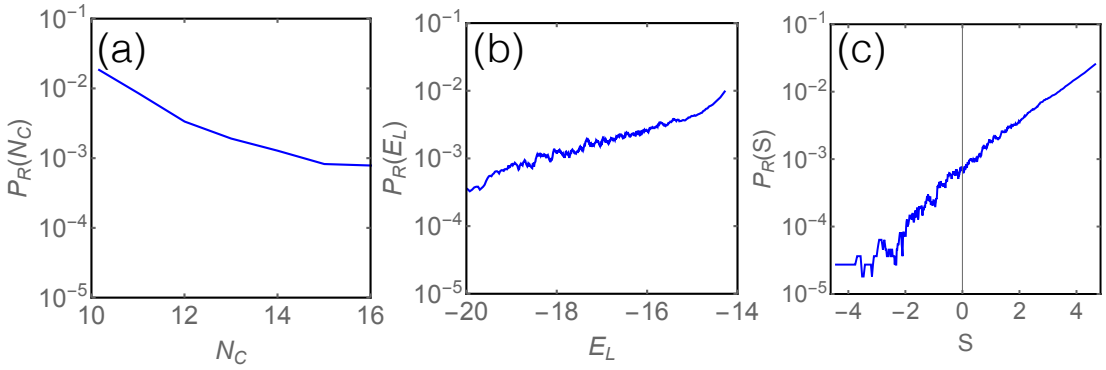


Figure 6.2: The probability of rearrangement vs. coordination number, local energy and softness. Average probabilities are calculated with a fixed bin size of 10^6 points.

A number of measures of local structure have been proposed to correlate with dynamics[181,192,196–201]. We have just shown that two of these quantities, local energy and coordination number, correlate with dynamics but are only weakly predictive of rearrangements. It is therefore important to compare and contrast different functions of local structure based on their ability to predict plastic activity *a priori* (Fig. 6.2). Here we introduce a metric to evaluate how well a given structural quantity associated with particle i, X_i, predicts rearrangements. In this description we assume that X correlates reasonably well with rearrangements, so that the probability that a particle rearranges, $P_R(X_i)$, is monotonically increasing in X_i. Note that there is a dis-

tribution of X_i values; we define its mean to be μ_X and its standard deviation to be σ_X.

The predictive power of a measure is related to how much more likely "high" X_i particles are to rearrange than their low X_i counterparts. We therefore propose the measure,

$$Q = \frac{P_R(X_i > \mu_X + \sigma_X)}{P_R(X_i < \mu_X - \sigma_X)} \tag{6.5}$$

which is the ratio of the probability of rearrangement for high X_i particles with that of low X_i particles. The higher the value of Q, the better X discriminates between particles likely to rearrange and those unlikely to rearrange, and therefore the better X is as a predictor of rearrangements. We find that for coordination number, $Q = 6.2$, and local energy, $Q = 5.0$, while for softness $Q = 165$. Thus, softness allows for two orders of magnitude more contrast between soft and hard particles than either of the other measures. We suggest that Q be used in future studies introducing new measures of local structure as predictors of rearrangements. To see this in practice, we estimate that for the method introduced in Yang *et al.*[196] $Q \approx 20$. We note that Q will be temperature dependent, and a fair comparison of different predictors should be done for systems under similar physical conditions. We can check the size of this temperature dependence by repeating the measurement of Q at $T = 0.35$. We find that $Q \gtrsim 500$ for softness while $Q = 7.6$ for coordination number and $Q = 9.9$ for local energy. Thus, while Q will in general be protocol dependent, it seems as though order of magnitude comparisons may still be made across different systems.

One might ask whether prediction accuracy improves if we use both local energy and coordination number information. We can characterize the local environment with two descriptors, the local potential energy and number of particles in the first

coordination shell, and calculate the line separating rearranging particles from non-rearranging particles in this two-dimensional space. When we do this, we find that the accuracy remains 65%. The fact that the prediction accuracy does not improve implies that the local energy and coordination number are highly correlated with one another so that no new information is added when both quantities are considered instead of just one of the two. The correlation of local energy with coordination number is shown in Fig. 6.1(b).

The situation is very different if two uncorrelated descriptors are used. As shown in Chapter 5.5, the best accuracy that can be achieved with one radial structure function (*i.e.* the value of the local pair correlation function $g(r)$ at one value of r) is 64%. However, if the best two structure functions are used, the accuracy increases to 77%. With only 10 radial structure functions, it is possible to reach 84% accuracy (Section 5.5). This is an example of a classic problem that arises in ML when there are correlations in the set of descriptors used. In physics, it is frequently the case that descriptors chosen for their physical significance are correlated with one another, and thus are likely to be biased in terms of the physical information they provide.

Finally, we study the relationship between softness and geometrical properties of the local neighborhoods. We first construct the Delaunay triangulation of the positions of the particles. Then we identify the tetrahedra that can be considered a regular tetrahedron or a regular quarter-octahedron, details of which are described in Ref.[69,202]. In Fig. 6.3, we plot the average softness of particles that belong to a corresponding number of regular tetrahedra and regular quarter-octahedra. The standard deviation of each point is comparable to the error bars in Fig 6.1. The average softness of a particle goes down if the tetrahedra it belongs to are regular. On the other hand, the average softness of a particle goes up with the number of good quarter-

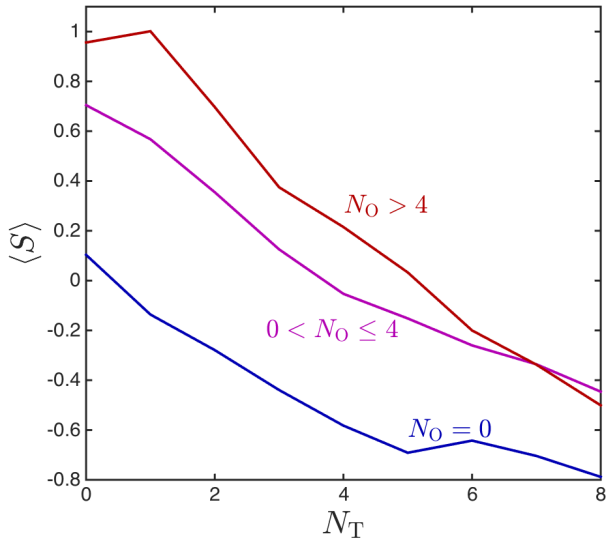


Figure 6.3: Average softness for particles that are on N_{T} regular tetrahedra. The averages are shown for particles that are on different numbers of regular quarter-octahedra N_{O}.

octahedra that it belongs to. The average softness of a particle that is on no good tetrahedra, but more than 4 good quarter-octahedra, is about 1. Particles that are not on any good quarter-octahedra, but on 8 good tetrahedra, have an average softness of -0.8. It would be interesting to consider the softness of particles in larger geometrical clusters, as they have been shown to be long-lived[181].

6.4 Predicting properties of rearrangements

In addition to predicting the probability of rearrangement, $P_R(S)$, softness contains additional predictive information about rearrangements. We first show that there is spatial structure in the softness field and that these spatial correlations appear to be related to the size of rearrangements in the system. Inspired by these results, we then show that spatial gradients in softness correlate with the direction of motion of parti-

cles during rearrangements.

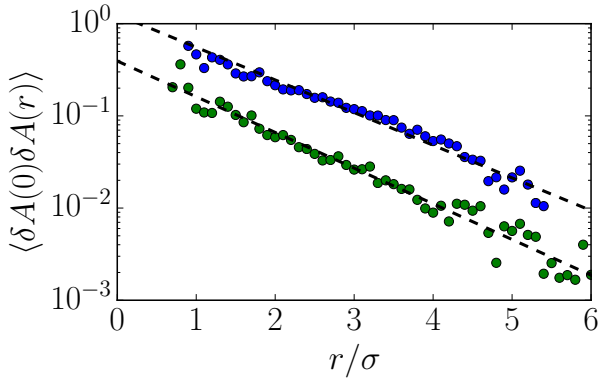


Figure 6.4: Here we plot the spatial correlation functions for softness (blue) and rearrangements (green). We overlay the best exponential fits with $\xi = 1.2$ for softness and $\xi = 1.1$ for rearrangements.

To investigate the spatial correlations of the softness field we consider normalized fluctuations in the field,

$$\delta S_i = \frac{S_i - \langle S \rangle}{\sqrt{\langle (S - \langle S \rangle)^2 \rangle}}. \tag{6.6}$$

We then compute the correlation function $\langle \delta S(0) \delta S(r) \rangle$ which is averaged over all particles separated by a distance r. This correlation function is shown in Fig. 6.4 in blue. We see that the softness field exhibits approximately exponential spatial correlations so that,

$$\langle \delta S(0) \delta S(r) \rangle = A e^{-r/\xi}. \tag{6.7}$$

Shown in dashed overlay is a best-fit to the correlation function with $\xi \approx 1.2$. Thus, our results show that the spatial extent of hard or soft regions is of order a single particle diameter. This length scale is commensurate with structural correlation lengths previously identified in glassy systems[203].

To measure the spatial extent of rearrangements we consider the spatial correlation of p_{hop}, $\langle \delta p(0) \delta p(r) \rangle$. As with softness we find that the correlations decay exponentially with $\xi \approx 1.1$. Thus, the characteristic size of localized rearrangements in the system is of order the particle diameter, reflecting the scale of spatial correlations of softness.

We next define a "gradient" of softness over the set of bonds in a neighborhood around a central particle i,

$$\nabla S_i = \sum_{j \text{ s.t. } R_{ij} < R_c} \frac{(S_j - S_i)\boldsymbol{R}_{ij}}{R_{ij}^2}. \tag{6.8}$$

We choose the neighborhoods of size $R_c = 2.5\sigma_{AA}$; however, our results are qualitatively insensitive to this choice.

For each rearrangement, α, involving a particle i we define t_α^{start} and t_α^{end} to be the start time and end time for the rearrangement respectively. We may then compute a displacement vector for the rearrangement,

$$\boldsymbol{\Delta r}_{i\alpha} = \boldsymbol{r}_i(t_\alpha^{\text{end}}) - \boldsymbol{r}_i(t_\alpha^{\text{start}}). \tag{6.9}$$

To show that rearrangements correlate with the gradient of the softness field we consider the cosine of the angle between the gradient of softness at a particle i right before it rearranges and the displacement vector of the particle during rearrangement,

$$\cos\theta = \frac{\boldsymbol{\Delta r} \cdot \boldsymbol{\nabla S}}{|\Delta r||\nabla S|}. \tag{6.10}$$

In addition to considering $\cos\theta$ we note that $|\Delta r||\nabla S|$ will be large when both the displacement during a rearrangement is large and the particle rearranges in a region

of high softness gradient. We focus on high values of $|\Delta r||\nabla S|$ since we would not expect very small displacements or rearrangements in regions of constant softness to have their motion described accurately by the spatial structure of the softness field.

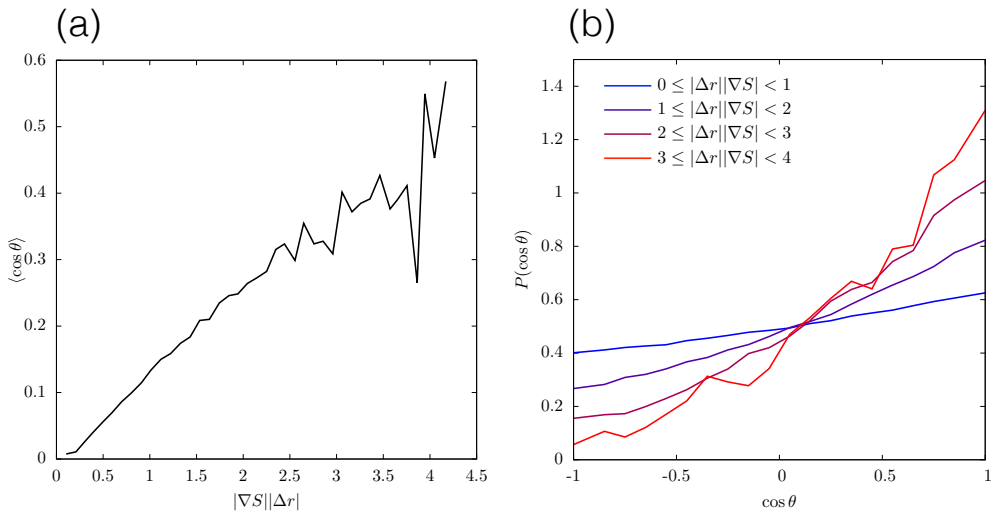


Figure 6.5: (a) The average angle between the rearrangement vector for a particle i, $\boldsymbol{\Delta r_i}$ and the gradient of its softness $\boldsymbol{\nabla S_i}$ as a function of the product of the norms $|\Delta r||\nabla S|$. We see that particles that exhibit significant rearrangements in a region of high gradient tend to move in the direction of the gradient. (b) The distribution, $P(\cos\theta)$, of the angle between rearrangements and the softness gradient at the rearranging particle. We plot different distributions for particles with different $|\Delta r||\nabla S|$ from low (blue) to high (red).

We plot in Fig. 6.5 (a) $\langle \cos\theta \rangle$ as a function of $|\Delta r||\nabla S|$. We find that $\langle \cos\theta \rangle$ is a strong function of $|\Delta r||\nabla S|$ increasing from 0 at small $|\Delta r||\nabla S|$ to about 0.45 at large $|\Delta r||\nabla S|$. At high $|\Delta r||\nabla S|$, we see that displacements tend to occur in the direction of increasing softness in a cone of $\approx$ 60 deg about $\boldsymbol{\nabla S}$. This result is supported in Fig. 6.5 (b) where we plot the distribution of $\cos\theta$ for different values of $|\Delta r||\nabla S|$. Strikingly we find that at large $|\Delta r||\nabla S|$ very few rearrangements tend to be in the direction of decreasing softness. These results may give insight into why rearrangements in supercooled liquids tend to be stringlike[204].

6.5 Simplifying softness

Here we use the fact that for the $d = 3$ Kob-Andersen model, it suffices to use only radial structure information. This allows us to recast softness in a form in which approximations become apparent that might lead the way to analytic progress on softness in the future. We start with a derivation for a single component system, noting that the extension to a system of several components (as in the KA mixture) is straightforward.

As described above, softness is defined in the context of a hyperplane, defined by a "weight" vector $\boldsymbol{w}$ and a bias b as well as a vector of structure functions $\boldsymbol{G}_i$ for each particle, i. In this context the softness is defined to be $S_i = \boldsymbol{w} \cdot \boldsymbol{G}_i + b$. We can explore the physical meaning of softness by considering the weights of the different structure functions, w_α. Structure functions that distinguish more effectively between rearranging and non-rearranging particles have larger projections onto the hyperplane, and therefore larger values of w. Additionally for the remainder of this discussion we will set the overall shift $b = 0$ since physical quantities such as the probability of rearrangement do not depend on it.

To begin suppose that we choose radial functions so that $G_{i\mu} = G(i; \mu\sigma, \sigma)$. Thus, the components of our feature vector are structure functions chosen such that the μth component is centered at a radius of $r_\mu = \mu\sigma$. If we would to consider the local structure out to a distance R_c then we will need at least $N_\sigma = R_c/\sigma$ structure functions. If we use the notation $w(r_\mu)$ for w_μ then we may express the softness as

$$S_i = \sum_{\mu=0}^{N_\sigma} w(r_\mu) G(i; r_\mu, \sigma). \tag{6.11}$$

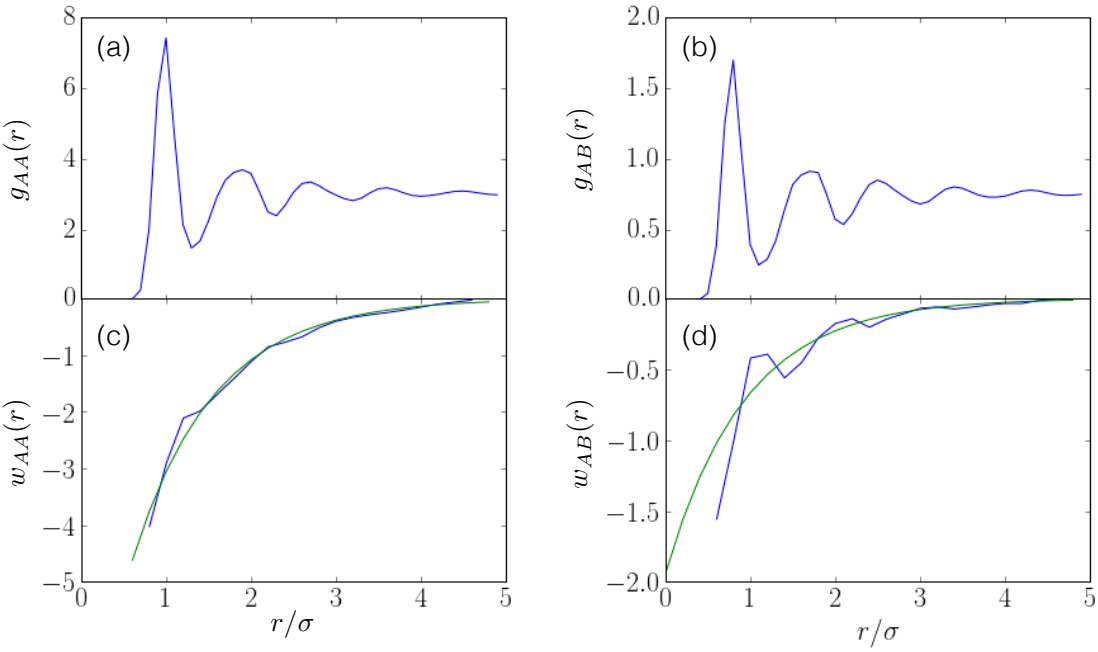


Figure 6.6: The weights as a function of radius. (a)-(b) the radial distribution functions for the three-dimensional Kob-Andersen system, $g_{AA}(r)$ and $g_{AB}(r)$ respectively. (c)-(d) the weights as a function of radius for these two pairs of species (blue) as well as exponential fits to these weights (green).

We can formally take the $\sigma \to 0$ limit, letting $r = r_\mu = \mu\sigma$ and note that

$$\lim_{\sigma\to 0} \sum_{\mu=0}^{N_\sigma} \sigma = \int_0^{R_c} dr \tag{6.12}$$

$$\lim_{\sigma\to 0} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2} r^2} = \delta(r) \tag{6.13}$$

to arrive at the expression,

$$S_i = \int_0^{R_c} dr \left[\sum_j \delta(R_{ij} - r) \right] w(r). \tag{6.14}$$

We note that the expression in braces is the local density of particles a distance r

away from a central particle i. It follows that we may regard softness as the functional of local density that best correlates with rearrangement dynamics. The approximation to the weight function that we obtain from ML is shown in Fig. 6.6. As an aside, note that eq.(6.14) implies that once the weight function is known, the softness may alternatively be computed as $S_i = \sum_j w(R_{ij})$. We have confirmed that this does indeed give a predictive accuracy that is identical to that of the usual method.

It is worthwhile to consider in more detail the form of the weight function and – in particular – approximations to the weight function that simplify analytic calculations. Fig. 6.6(c)-(d) shows the weight functions for species A particles from the 3D KA mixture. Note that since there are now two species each species will have two weight functions, one for each type of interaction. We first note that the weight functions are well-approximated by the exponential form. We show exponential fits to the weight functions in green in Fig. 6.6(c)-(d). Using this approximation we find that we are able to successfully predict 82.5% of rearrangements (down from 88% using the full weight function.) Although we have lost some accuracy in this approach we note that the softness calculation simply becomes

$$S_i = \sum_{j \in A} A_{AA} e^{-\alpha_{AA} R_{ij}} + \sum_{j \in B} A_{AB} e^{-\alpha_{AB} R_{ij}}. \tag{6.15}$$

However, we note that the other property of the weight function is that it is positive for $r < \sigma^{HS}$ where σ^{HS} is the effective hard-sphere radius. Thus, we try the slightly more physical approximation,

$$w_{XY}(r) = \begin{cases} C_{XY} & r < \sigma_{XY}^{HS} \\ A_{XY} e^{-\alpha_{XY} r} & r > \sigma_{XY}^{HS}. \end{cases} \tag{6.16}$$

Using this approximation we are able to predict 85% of the rearrangements correctly. Thus, we see that we may use relatively simple approximations to the rather complicated weight functions while retaining strong correlations with dynamics. These approximations have the added benefit that they are motivated by data as opposed to more ad-hoc approximations such as local potential energy and free-volume which correlate much less well with dynamics.

6.6 Unsupervised dimensionality reduction

The support vector machines method used to construct softness, as described above, is a "supervised learning" technique. Supervised learning requires a training set, which consists of descriptors (the structure functions calculated for each particle), and labels (whether a particle is about to rearrange, or whether it has not rearranged for a long time).

For unsupervised ML techniques, labels are not needed. The assumption is that there is a true structural difference between particles that are about to rearrange and those that are not, which could be discerned *a priori*. In unsupervised learning, the ML model is instead applied directly to the structure functions that make up the training set. We use the t-distributed stochastic neighbor embedding (t-SNE) method, which is a state-of-the-art technique for dimensionality reduction, on the M-dimensional structure function space, $\mathbb{R}^M$. t-SNE has been applied successfully to visualizing complex data sets, e.g. handwritten numbers[134], and game states of neural networks as they play Atari games[205]. It projects a high dimensional data set onto two or three dimensions by trying to place points that are similar to each other in the high dimensional space closer to each other in the low dimensional projection. t-SNE is a non-parametric mapping, which means that the dimensions in the reduced space

are not explicit functions of the dimensions in the original high-dimensional space. This prevents us from interpreting the physical significance of the reduced dimensions. However, similar local structures are mapped onto points that are closer together in the low-dimensional space; this allows us to infer the properties of the data that correspond to softness.

We apply t-SNE, using the Barnes-Hut approximation[206,207], to data from the two different systems studied: the $d = 2$ LJ supercooled liquid under shear, and $d = 3$ LJ supercooled liquid in equilibrium. Each data set consists of 16,000 particles that are about to rearrange and 16,000 particles that are not. Without the supervision of labels, the results are sensitive to the selection of descriptors. For this reason, we excluded the first 6 radial structure functions. These excluded functions describe the neighbors for a particle that are closer than $0.3\sigma_{AA}$. We use all the remaining descriptors (94 radial and 66 angular structure functions) for dimensionality reduction. The resulting two dimensional maps are plotted in Fig. 6.7. We color points red if they correspond to particles about to rearrange, and blue if they correspond to particles that are not. These labels are *not* used in the application of the t-SNE algorithm; they are applied afterwards, only for visualization purposes. For both systems, there is a clear separation of red points from blue points, indicating that there is truly a structural difference between particles that are likely to rearrange and those that are not. We have also applied the force directed graph drawing algorithm[208] on the same data, in which springs are introduced between points in the high-dimensional space and the total energy stored in the springs is minimized as the system is collapsed onto the low-dimensional space. This approach leads to the same qualitative results.

Since t-SNE and force directed graph drawing algorithm are non-parametric methods, it is difficult to interpret how the projection depends on different structure func-

tions (whereas after the application of SVM, one can compare the weights corresponding to each structure function). We observed that the radial structure functions from $0.3\sigma_{AA}$ up to $2.5\sigma_{AA}$ and all of the angular functions are required to obtain the results reported here.

Note also that the sheared $d = 2$ system shows two distinct populations of particles (blue) that are not rearranging. This difference between these two systems was previously observed; the SVM method also shows that there are two structurally distinct populations of hard particles in the $d = 2$ sheared LJ liquid. By contrast, the $d = 3$ system seems to have only one population of rearranging and and one population of non-rearranging particles. We speculate that the difference between $d = 2$ and $d = 3$ arises from the relatively few options for local packing geometry in $d = 2$.

6.7 Conclusion

We have investigated different methods for elucidating the relationship between structure and dynamics in disordered solids, using $d = 2$ and $d = 3$ LJ supercooled liquids as model systems. We find that while many easily-calculated structural quantities, such as local potential energy and coordination number, correlate strongly with softness and dynamics, they can only predict rearrangements with an accuracy of 60-65%, even when used in tandem. By contrast, support vector machines allow us to predict rearrangements using a function of 160 different radial and angular structure functions with an accuracy of 90%. We showed that for the LJ systems studied, simpler models with only 8 parameters can be used with only a slight loss of accuracy (85% instead of 90%).

We recast softness to pave the way for future analytical approaches to glassy dynamics based on structure. Finally, we show that it is possible to capture certain as-

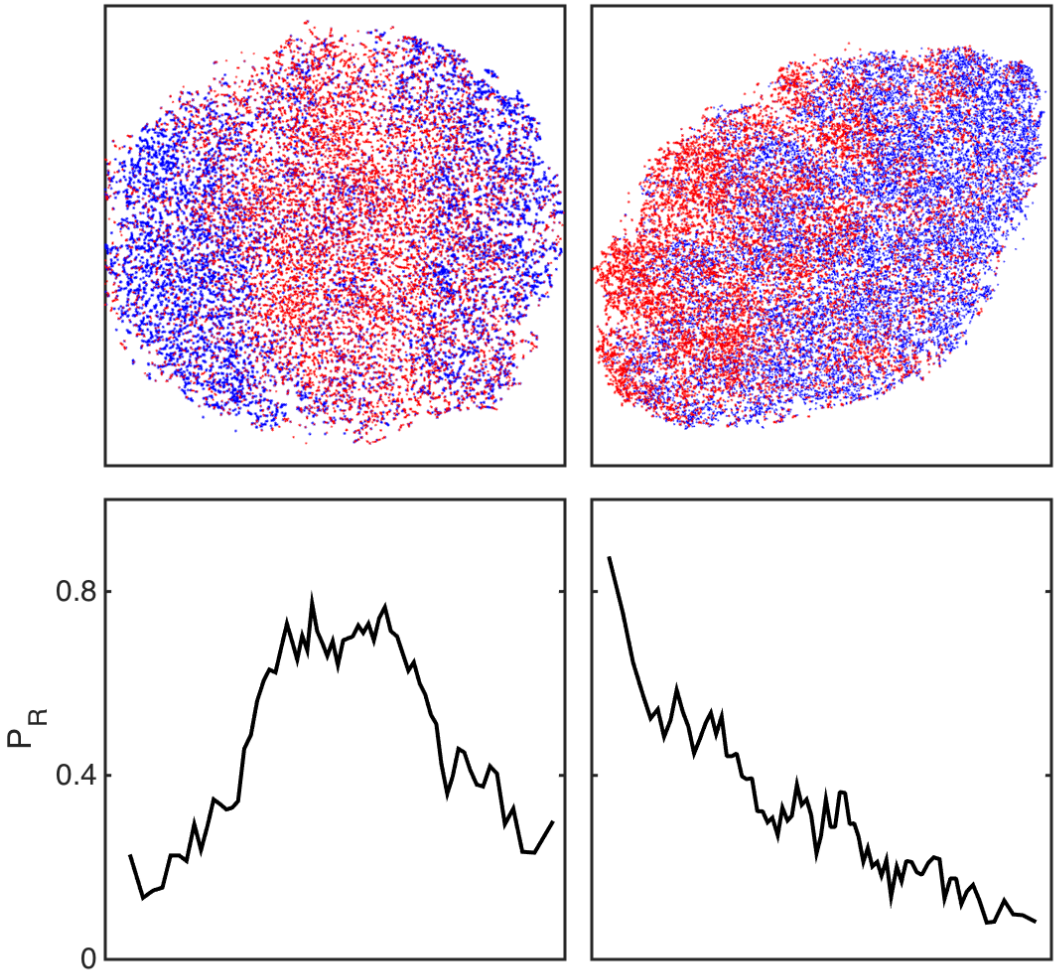


Figure 6.7: Unsupervised learning on d=2 (left) and d=3 (right) Lennard Jones liquids: top panels represent the structure functions after dimensionality reduction. Particles that are about to rearrange or not about to rearrange are colored red and blue, respectively. Bottom panels represent the probability of rearrangement, for bins of fixed size (500 points) spanning the two-dimensional spaces shown in the upper panels. Red particles are localized to the center in the top left panel; these particles have a higher probability of rearranging (bottom left). Similarly, the probability of a rearrangement is clearly different for different regions of the two-dimensional space.

pects of soft particles using unsupervised learning, where a labelled training set is not required. However, these individual structural quantities and the unsupervised models

do not have predictive power, as they cannot capture the complexity of patterns in disordered LJ structures.

References

[1] A. S. Keys, L. O. Hedges, J. P. Garrahan, S. C. Glotzer, and D. Chandler. Excitations are localized and relaxation is hierarchical in glass-forming liquids. *Phys. Rev. X.*, 1:021013, 2011.

[2] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *Apl Materials*, 1(1):011002, 2013.

[3] Gerbrand Ceder. Opportunities and challenges for first-principles materials design and applications to li battery materials. *MRS bulletin*, 35(09):693–701, 2010.

[4] Thomas W Eagar. Bringing new materials to market. *Technology Review*, 98(2):42–49, 1995.

[5] JP Holdren et al. Materials genome initiative: Strategic plan. *Washington DC: Office of Science and Technology Policy*, 6, 2014.

[6] G Ceder and K Persson. How supercomputers will yield a golden age of materials science. *Scientific American, Dec*, 2013.

[7] Edward O Pyzer-Knapp, Changwon Suh, Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, and Alán Aspuru-Guzik. What is high-throughput virtual screening? a perspective from organic materials discovery. *Annual Review of Materials Research*, 45:195–216, 2015.

[8] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[9] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[10] John P Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048, 1981.

[11] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[12] John P Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B*, 45(23):13244, 1992.

[13] Axel D Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of chemical physics*, 98(7):5648–5652, 1993.

[14] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.

[15] A Daniel Boese and Nicholas C Handy. A new parametrization of exchange–correlation generalized gradient approximation functionals. *The Journal of Chemical Physics*, 114(13):5497–5503, 2001.

[16] Yan Zhao and Donald G Truhlar. Density functionals with broad applicability in chemistry. *Accounts of chemical research*, 41(2):157–167, 2008.

[17] DR Hartree and W Hartree. Self-consistent field with exchange for potassium and argon. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 450–464, 1938.

[18] V Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörper-problems. *Zeitschrift für Physik*, 61(1-2):126–148, 1930.

[19] John C Slater. A simplification of the hartree-fock method. *Physical Review*, 81(3):385, 1951.

[20] Lars Hedin. New method for calculating the one-particle green's function with application to the electron-gas problem. *Physical Review*, 139(3A):A796, 1965.

[21] L Hedin and A Johansson. Polarization corrections to core levels. *Journal of Physics B: Atomic and Molecular Physics*, 2(12):1336, 1969.

[22] Mark S Hybertsen and Steven G Louie. First-principles theory of quasiparticles: calculation of band gaps in semiconductors and insulators. *Physical review letters*, 55(13):1418, 1985.

[23] Mark S Hybertsen and Steven G Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B*, 34(8):5390, 1986.

[24] Giovanni Onida, Lucia Reining, and Angel Rubio. Electronic excitations: density-functional versus many-body green's-function approaches. *Reviews of Modern Physics*, 74(2):601, 2002.

[25] WMC Foulkes, Lubos Mitas, RJ Needs, and G Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33, 2001.

[26] B Tanatar and DM Ceperley. Ground state of the two-dimensional electron gas. *Physical Review B*, 39(8):5005, 1989.

[27] David M Ceperley and BJ Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566, 1980.

[28] Peter J Reynolds, David M Ceperley, Berni J Alder, and William A Lester Jr. Fixed-node quantum monte carlo for moleculesa) b). *The Journal of Chemical Physics*, 77(11):5593–5603, 1982.

[29] John C Slater and George F Koster. Simplified lcao method for the periodic potential problem. *Physical Review*, 94(6):1498, 1954.

[30] Ole Krogh Andersen and Ove Jepsen. Explicit, first-principles tight-binding theory. *Physical Review Letters*, 53(27):2571, 1984.

[31] Shiang Fang, Rodrick Kuate Defo, Sharmila N Shirodkar, Simon Lieu, Georgios A Tritsaris, and Efthimios Kaxiras. Ab initio tight-binding hamiltonian for transition metal dichalcogenides. *Physical Review B*, 92(20):205108, 2015.

[32] John Edward Jones. On the determination of molecular fields. ii. from the equation of state of a gas. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 106, pages 463–477. The Royal Society, 1924.

[33] Frank H Stillinger and Thomas A Weber. Computer simulation of local order in condensed phases of silicon. *Physical review B*, 31(8):5262, 1985.

[34] J Tersoff. New empirical approach for the structure and energy of covalent systems. *Physical Review B*, 37(12):6991, 1988.

[35] J Tersoff. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B*, 39(8):5566, 1989.

[36] Murray S Daw and Michael I Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical Review B*, 29(12):6443, 1984.

[37] Murray S Daw, Stephen M Foiles, and Michael I Baskes. The embedded-atom method: a review of theory and applications. *Materials Science Reports*, 9(7):251–310, 1993.

[38] Martin Z Bazant, Efthimios Kaxiras, and JF Justo. Environment-dependent interatomic potential for bulk silicon. *Physical Review B*, 56(14):8542, 1997.

[39] Adri CT Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. Reaxff: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

[40] Kimberly Chenoweth, Adri CT Van Duin, and William A Goddard. Reaxff reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *The Journal of Physical Chemistry A*, 112(5):1040–1053, 2008.

[41] Gus LW Hart, Volker Blum, Michael J Walorski, and Alex Zunger. Evolutionary approach for determining first-principles hamiltonians. *Nature materials*, 4(5):391–394, 2005.

[42] Volker Blum, Gus LW Hart, Michael J Walorski, and Alex Zunger. Using genetic algorithms to map first-principles results to model hamiltonians: Application to the generalized ising model for alloys. *Physical Review B*, 72(16):165113, 2005.

[43] M. L. Manning and A. J. Liu. *Phys. Rev. Lett.*, 107:108302, 2011.

[44] Srikanth Sastry, Pablo G Debenedetti, and Frank H Stillinger. Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid. *Nature*, 393(6685):554–557, 1998.

[45] Pablo G Debenedetti and Frank H Stillinger. Supercooled liquids and the glass transition. *Nature*, 410(6825):259–267, 2001.

[46] Giorgio Parisi and Francesco Zamponi. Mean-field theory of hard sphere glasses and jamming. *Rev. Mod. Phys.*, 82(789–845):789, 2010.

[47] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Exact theory of dense amorphous hard spheres in high dimensions. iii. the full rsb solution. *J. Stat. Mech.*, page 10009, 2014.

[48] Sadanand Singh, MD Ediger, and Juan J de Pablo. Ultrastable glasses from in silico vapour deposition. *Nature materials*, 12(2):139–144, 2013.

[49] Hai-Bin Yu, Yuansu Luo, and Konrad Samwer. Ultrastable metallic glass. *Advanced Materials*, 25(41):5904–5908, 2013.

[50] Chris J Pickard and RJ Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011.

[51] Artem R Oganov and Colin W Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of chemical physics*, 124(24):244704, 2006.

[52] David J Wales and Harold A Scheraga. Global optimization of clusters, crystals, and biomolecules. *Science*, 285(5432):1368–1372, 1999.

[53] Paul AM Dirac. Quantum mechanics of many-electron systems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 123, pages 714–733. The Royal Society, 1929.

[54] Llewellyn H Thomas. The calculation of atomic fields. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 23, pages 542–548. Cambridge Univ Press, 1927.

[55] Enrico Fermi. Un metodo statistico per la determinazione di alcune priorieta dell'atome. *Rend. Accad. Naz. Lincei*, 6(602-607):32, 1927.

[56] Enrico Fermi. Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente. *Zeitschrift für Physik*, 48(1-2):73–79, 1928.

[57] Paul AM Dirac. Note on exchange phenomena in the thomas atom. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 26, pages 376–385. Cambridge Univ Press, 1930.

[58] Robert O Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of modern physics*, 87(3):897, 2015.

[59] Efthimios Kaxiras. *Atomic and electronic structure of solids*. Cambridge University Press, 2003.

[60] Erich Strohmaier, Hans W Meuer, Jack Dongarra, and Horst D Simon. The top500 list and progress in high-performance computing. *Computer*, (11):42–49, 2015.

[61] Lieven MK Vandersypen, Matthias Steffen, Gregory Breyta, Costantino S Yannoni, Mark H Sherwood, and Isaac L Chuang. Experimental realization of shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature*, 414(6866):883–887, 2001.

[62] Zhengbing Bian, Fabian Chudak, William G Macready, Lane Clark, and Frank Gaitan. Experimental determination of ramsey numbers. *Physical review letters*, 111(13):130505, 2013.

[63] Stefano Curtarolo, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kesong Yang, Richard H Taylor, Lance J Nelson, Gus LW Hart, Stefano Sanvito, Marco Buongiorno-Nardelli, et al. Aflowlib. org: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science*, 58:227–235, 2012.

[64] Johannes Hachmann, Roberto Olivares-Amaya, Sule Atahan-Evrenk, Carlos Amador-Bedolla, Roel S Sánchez-Carrera, Aryeh Gold-Parker, Leslie Vogt, Anna M Brockway, and Alán Aspuru-Guzik. The harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the

world community grid. *The Journal of Physical Chemistry Letters*, 2(17):2241–2251, 2011.

[65] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98(14):146401, 2007.

[66] Albert P Bartók, Mike C Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104(13):136403, 2010.

[67] Atsuto Seko, Akira Takahashi, and Isao Tanaka. Sparse representation for a potential energy surface. *Phys. Rev. B*, 90(2):024101, 2014.

[68] John C Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding density functionals with machine learning. *Physical review letters*, 108(25):253002, 2012.

[69] Ekin D Cubuk and Efthimios Kaxiras. Theory of structural transformation in lithiated amorphous silicon. *Nano Lett.*, 14(7):4065–4070, 2014.

[70] Ekin Dogus Cubuk, Samuel Stern Schoenholz, Jennifer M. Rieser, Brad Dean Malone, Joerg Rottler, Douglas J. Durian, Efthimios Kaxiras, and Andrea J. Liu. Identifying structural flow defects in disordered solids using machine learning methods. *Phys. Rev. Lett.*, 114(10):108001, 2015.

[71] Samuel Stern Schoenholz, Ekin Dogus Cubuk, Daniel M. Sussman, Efthimios Kaxiras, and Andrea J. Liu. *Nat. Phys.*, 2016.

[72] Ekin Dogus Cubuk, Samuel Stern Schoenholz, Efthimios Kaxiras, and Andrea J Liu. Structural properties of defects in glassy liquids. *J. Phys. Chem.*, 2016.

[73] M. Armand and J. M. Tarascon. Building better batteries. *Nature*, 451(7179):652–657, 2008.

[74] M. S. Whittingham. Materials challenges facing electrical energy storage. *MRS Bull.*, 33(4):411–419, 2008.

[75] U. Kasavajjula, C. S. Wang, and A. J. Appleby. Nano- and bulk-silicon-based insertion anodes for lithium-ion secondary cells. *J. Power Sources*, 163(2):1003–1039, 2007.

[76] W. J. Zhang. A review of the electrochemical performance of alloy anodes for lithium-ion batteries. *J. Power Sources*, 196(1):13–24, 2011.

[77] C. K. Chan, H. L. Peng, G. Liu, K. McIlwrath, X. F. Zhang, R. A. Huggins, and Y. Cui. High-performance lithium battery anodes using silicon nanowires. *Nat. Nanotechnol.*, 3(1):31–35, 2008.

[78] C. J. Wen and R. A. Huggins. Chemical diffusion in intermediate phases in the lithium-silicon system. *J. Solid State Chem.*, 37(3):271–278, 1981.

[79] L. Y. Beaulieu, T. D. Hatchard, A. Bonakdarpour, M. D. Fleischauer, and J. R. Dahn. Reaction of li with alloy thin films studied by in situ afm. *J. Electrochem. Soc.*, 150(11):A1457–A1464, 2003.

[80] Matthew T McDowell, Seok Woo Lee, Justin T Harris, Brian A Korgel, Chongmin Wang, William D Nix, and Yi Cui. In situ tem of two-phase lithiation of amorphous silicon nanospheres. *Nano Lett.*, 13(2):758–764, 2013.

[81] Ekin D Cubuk, Wei L Wang, Kejie Zhao, Joost J Vlassak, Zhigang Suo, and Efthimios Kaxiras. Morphological evolution of si nanowires upon lithiation: A first-principles multiscale model. *Nano Lett.*, 13(5):2011–2015, 2013.

[82] Maria K. Y. Chan, C. Wolverton, and Jeffrey P. Greeley. First principles simulations of the electrochemical lithiation and delithiation of faceted crystalline silicon. *J. Am. Chem. Soc.*, 134(35):14362–14374, 2012.

[83] Sung Chul Jung, Jang Wook Choi, and Young-Kyu Han. Anisotropic volume expansion of crystalline silicon during electrochemical lithium insertion: An atomic level rationale. *Nano Lett.*, 12(10):5342–5347, 2012.

[84] M. J. Chon, V. A. Sethuraman, A. McCormick, V. Srinivasan, and P. R. Guduru. Real-time measurement of stress and damage evolution during initial lithiation of crystalline silicon. *Phys. Rev. Lett.*, 107:045503, Jul 2011.

[85] S.W. Lee, L.A. Berla, M.T. McDowell, W.D. Nix, and Y. Cui. Reaction front evolution during electrochemical lithiation of crystalline silicon nanopillars. *Isr. J. Chem.*, 52:1118–1123, 2012.

[86] S.W. Lee, M.T. McDowell, L.A. Berla, W.D. Nix, and Y. Cui. Fracture of crystalline silicon nanopillars during electrochemical lithium insertion. *Proc. Natl. Acad. Sci.*, 109(11):4080–4085, 2012.

[87] X.H. Liu, J.W. Wang, S. Huang, F. Fan, X. Huang, Y. Liu, S. Krylyuk, J. Yoo, S.A. Dayeh, Mao S.X. Davydov, A.V., S.T. Picraux, S. Zhang, J. Li, T. Zhu, and Huang J.Y. In situ atomic-scale imaging of electrochemical lithiation in silicon. *Nat. Nanotechnol.*, 7:749–756, 2012.

[88] Matt Pharr, Kejie Zhao, Xinwei Wang, Zhigang Suo, and Joost J. Vlassak. Kinetics of initial lithiation of crystalline silicon electrodes of lithium-ion batteries. *Nano Lett.*, 12(9):5039–5047, 2012.

[89] M.T. McDowell, I. Ryu, S.W. Lee, C. Wang, W.D. Nix, and Y. Cui. Studying the kinetics of crystalline silicon nanoparticle lithiation with in situ transmission electron microscopy. *Adv. Mater.*, 24:6034–6041, 2012.

[90] Qianfan Zhang, Wenxing Zhang, Wenhui Wan, Yi Cui, and Enge Wang. Lithium insertion in silicon nanowires: An ab initio study. *Nano Lett.*, 10(9):3243–3249, 2010.

[91] K. J. Zhao, W. L. Wang, J. Gregoire, M. Pharr, Z. G. Suo, J. J. Vlassak, and E. Kaxiras. Lithium-assisted plastic deformation of silicon electrodes in lithium-ion batteries: A first-principles theoretical study. *Nano Lett.*, 11(7):2962–2967, 2011.

[92] Jiang Wei Wang, Yu He, Feifei Fan, Xiao Hua Liu, Shuman Xia, Yang Liu, C. Thomas Harris, Hong Li, Jian Yu Huang, Scott X. Mao, and Ting Zhu. Two-phase electrochemical lithiation in amorphous silicon. *Nano Lett.*, 13(2):709–715, 2013.

[93] V L Chevrier and J R Dahn. First principles model of amorphous silicon lithiation. *J. Electrochem. Soc.*, 156(6):A454–A458, 2009.

[94] Hyunwoo Kim, Chia-Yun Chou, John G Ekerdt, and Gyeong S Hwang. Structure and properties of li- si alloys: A first-principles study. *J. Phys. Chem. C*, 115(5):2514–2521, 2011.

[95] Shan Huang and Ting Zhu. Atomistic mechanisms of lithium insertion in amorphous silicon. *J. Power Sources*, 196(7):3664–3668, 2011.

[96] Zhiwei Cui, Feng Gao, Zhihua Cui, and Jianmin Qu. A second nearest-neighbor embedded atom method interatomic potential for li–si alloys. *J. Power Sources*, 207:150–159, 2012.

[97] Feifei Fan, Shan Huang, Hui Yang, Muralikrishna Raju, Dibakar Datta, Vivek B Shenoy, Adri C T van Duin, Sulin Zhang, and Ting Zhu. Mechanical properties of amorphous li x si alloys: a reactive force field study. *Modell. Simul. Mater. Sci. Eng.*, 21(7):074002, 2013.

[98] Alireza Ostadhossein, Ekin D Cubuk, Georgios A Tritsaris, Efthimios Kaxiras, Sulin Zhang, and Adri CT van Duin. Stress effects on the initial lithiation of crystalline silicon nanowires: reactive molecular dynamics simulations using reaxff. *Physical Chemistry Chemical Physics*, 17(5):3832–3840, 2015.

[99] Alireza Ostadhossein, Sung-Yup Kim, Ekin Dogus Cubuk, Yue Qi, and Adri CT van Duin. Atomic insight into the lithium storage and diffusion mechanism of sio2/al2o3 electrodes of li-ion batteries: Reaxff reactive force field modeling. *The Journal of Physical Chemistry A*, 2016.

[100] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal. The siesta method for ab initio order-n materials simulation. *J Phys.: Condens. Matter*, 14(11):2745–2779, 2002.

[101] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865, 1996.

[102] Martin Z Bazant, Efthimios Kaxiras, and J F Justo. Environment-dependent interatomic potential for bulk silicon. *Phys. Rev. B*, 56(14):8542, 1997.

[103] Yina Mo, Martin Z Bazant, and Efthimios Kaxiras. Sulfur point defects in crystalline and amorphous silicon. *Phys. Rev. B*, 70(20):205210, 2004.

[104] V L Chevrier and J R Dahn. First principles studies of disordered lithiated silicon. *J. Electrochem. Soc.*, 157(4):A392–A398, 2010.

[105] D S Franzblau. Computation of ring statistics for network models of solids. *Phys. Rev. B*, 44(10):4925, 1991.

[106] J D Joannopoulos and Marvin L Cohen. Theory of short-range order and disorder in tetrahedrally bonded semiconductors. *Solid State Phys.*, 31:71–148, 1976.

[107] N N Medvedev, Yu I Naberukhin, and V A Luchnikov. Perfect structure regions in amorphous argon. *J. Struct. Chem*, 35(1):47–56, 1994.

[108] N N Medvedev. Aggregation of tetrahedral and quartoctahedral delaunay simplices in liquid and amorphous rubidium. *J. Phys.: Condens. Matter*, 2(46):9145, 1990.

[109] Nikolai N Medvedev, Alfons Geiger, and Witold Brostow. Distinguishing liquids from amorphous solids: Percolation analysis on the voronoi network. *J. Chem. Phys.*, 93:8337, 1990.

[110] V A Luchnikov, N N Medvedev, A Appelhagen, and A Geiger. Medium-range structure of amorphous silicon studied by the voronoi¿delaunay method. *Mol. Phys.*, 88(5):1337–1348, 1996.

[111] Alexey V Anikeenko, Marina L Gavrilova, and Nikolai N Medvedev. A novel delaunay simplex technique for detection of crystalline nuclei in dense packings of spheres. Lect. Notes Comput. Sc., pages 816–826. Springer, 2005.

[112] M N Obrovac and Leif Christensen. Structural changes in silicon anodes during lithium insertion/extraction. *Electrochem. Solid-State Lett.*, 7(5):A93–A96, 2004.

[113] Meng Gu, Zhiguo Wang, Justin G Connell, Daniel E Perea, Lincoln J Lauhon, Fei Gao, and Chongmin Wang. Electronic origin for the phase transition from amorphous li x si to crystalline li15si4. *ACS Nano*, 7(7):6303–6309, 2013.

[114] Yoshiyuki Kubota, Mary Clare Sison Escaño, Hiroshi Nakanishi, and Hideaki Kasai. Crystal and electronic structure of li15si4. *J. Appl. Phys.*, 102(5):053704, 2007.

[115] RA Johnson and DJ Oh. Analytic embedded atom method model for bcc metals. *J. Mater. Res.*, 4(05):1195–1201, 1989.

[116] Gabor Csányi, T Albaret, MC Payne, and A De Vita. learn on the fly : a hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.*, 93(17):175503, 2004.

[117] Zhenwei Li, James R Kermode, and Alessandro De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114(9):096405, 2015.

[118] Amos Waterland, Elaine Angelino, Ekin D Cubuk, Efthimios Kaxiras, Ryan P Adams, Jonathan Appavoo, and Margo Seltzer. Computational caches. In *Proceedings of the 6th International Systems and Storage Conference*, page 8. ACM, 2013.

[119] Danny Perez, Ekin D. Cubuk, Amos Waterland, Efthimios Kaxiras, and Arthur F. Voter. Long-time dynamics through parallel trajectory splicing. *J. Chem. Theory Comput.*, 12(1):18–28, 2016.

[120] Rustam Z Khaliullin, Hagai Eshet, Thomas D Kühne, Jörg Behler, and Michele Parrinello. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nat. Mater.*, 10(9):693–697, 2011.

[121] Jörg Behler, Roman Martoňák, Davide Donadio, and Michele Parrinello. Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys. Rev. Lett.*, 100(18):185501, 2008.

[122] Nongnuch Artrith and Jörg Behler. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B*, 85(4):045439, 2012.

[123] Nongnuch Artrith and Alexie M Kolpak. Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: A combination of dft and accurate neural network potentials. *Nano Lett.*, 14(5):2670–2676, 2014.

[124] Suresh Kondati Natarajan, Tobias Morawietz, and Jörg Behler. Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials. *Phys. Chem. Chem. Phys.*, 2015.

[125] Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *J. Comput. Chem.*, 29(13):2044–2078, 2008.

[126] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47(1):558, 1993.

[127] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, 6(1):15–50, 1996.

[128] Nongnuch Artrith, Björn Hiller, and Jörg Behler. Neural network potentials for metals and oxides–first applications to copper clusters at zinc oxide. *Phys. Status Solidi (b)*, 250(6):1191–1203, 2013.

[129] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798–1828, 2013.

[130] Ronan Collobert, Koray Kavukcuoglu, and Clément Farabet. Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*, number EPFL-CONF-192376, 2011.

[131] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometr. Intell. Lab.*, 2(1):37–52, 1987.

[132] VA Luchnikov, NN Medvedev, A Appelhagen, and A Geiger. Medium-range structure of amorphous silicon studied by the voronoi delaunay method. *Mol. Phys.*, 88(5):1337–1348, 1996.

[133] Brad D Malone, Jay D Sau, and Marvin L Cohen. Ab initio survey of the electronic structure of tetrahedrally bonded phases of silicon. *Phys. Rev. B*, 78(3):035210, 2008.

[134] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(2579-2605):85, 2008.

[135] André Hedler, Siegfried Ludwig Klaumünzer, and Werner Wesch. Amorphous silicon exhibits a glass transition. *Nat. Mater.*, 3(11):804–809, 2004.

[136] David J Wales. Decoding the energy landscape: extracting structure, dynamics and thermodynamics. *Phil. Trans. R. Soc. A*, 370(1969):2877–2899, 2012.

[137] David J Wales, Mark A Miller, and Tiffany R Walsh. Archetypal energy landscapes. *Nature*, 394(6695):758–760, 1998.

[138] John C Jamieson. Crystal structures at high pressures of metallic modifications of silicon and germanium. *Science*, 139(3556):762–764, 1963.

[139] JZ Hu and IL Spain. Phases of silicon at high pressure. *Solid State Commun.*, 51(5):263–266, 1984.

[140] Bernd G Pfrommer, Michel Côté, Steven G Louie, and Marvin L Cohen. Ab initio study of silicon in the r 8 phase. *Phys. Rev. B*, 56(11):6662, 1997.

[141] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108(5):058301, 2012.

[142] Koji Fujimura, Atsuto Seko, Yukinori Koyama, Akihide Kuwabara, Ippei Kishida, Kazuki Shitara, Craig AJ Fisher, Hiroki Moriwake, and Isao Tanaka. Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms. *Advanced Energy Materials*, 3(8):980–985, 2013.

[143] Stefano Curtarolo, Gus LW Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito, and Ohad Levy. The high-throughput highway to computational materials design. *Nat. Mater.*, 12(3):191–201, 2013.

[144] Christopher C Fischer, Kevin J Tibbetts, Dane Morgan, and Gerbrand Ceder. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.*, 5(8):641–646, 2006.

[145] G. I. Taylor. The mechanism of plastic deformation of crystals. part i. theoretical. *Proc. R. Soc. Lond. A Math. Phys. Sci.*, 145:362, 1934.

[146] A. M Alsayed, M. F. Islam, J. Zhang, P. J. Collings, and A. G. Yodh. *Science*, 309:1207, 2005.

[147] A. S. Argon and H. Y. Kuo. Plastic flow in a disordered bubble raft (an analog of a metallic glass). *Materials science and Engineering*, 39(1):101–109, 1979.

[148] A. S. Argon. Plastic deformation in metallic glasses. *Acta Mater.*, 27(1):47–58, 1979.

[149] M. L. Falk and J. S. Langer. Dynamics of viscoplastic deformation in amorphous solids. *Phys. Rev. E*, 57(6):7192, 1998.

[150] L. Bocquet, A. Colin, and A. Ajdari. Kinetic theory of plastic flow in soft glassy materials. *Phys. Rev. Lett.*, 103(3):036001, 2009.

[151] Jörg Rottler, Samuel S. Schoenholz, and Andrea J. Liu. Predicting plasticity with soft vibrational modes: From dislocations to glasses. *Phys. Rev. E*, 89:042304, Apr 2014.

[152] C. Brito and M. Wyart. *J. Stat. Mech.*, page L08003, 2007.

[153] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman. *Nat. Phys.*, 4:711, 2008.

[154] Asaph Widmer-Cooper, Heidi Perry, Peter Harrowell, David R Reichman, et al. Localized soft modes and the supercooled liquid's irreversible passage through its configuration space. *J. Chem. Phys.*, 131(19):194508, 2009.

[155] R. Candelier, A. Widmer-Cooper, J. K. Kummerfeld, O. Dauchot, G. Biroli, P. Harrowell, and D. R. Reichman. Spatiotemporal hierarchy of relaxation events, dynamical heterogeneities, and structural reorganization in a supercooled liquid. *Phys. Rev. Lett.*, 105:135702, 2010.

[156] K. Chen, M. L. Manning, P. J. Yunker, W. G. Ellenbroek, Z. Zhang, A. J. Liu, and A. G. Yodh. *Phys. Rev. Lett.*, 107:108301, 2011.

[157] G. M. Hocky and D. R. Reichman. *J. Chem. Phys.*, 138:12A537, 2013.

[158] K. Chen, T. Still, S. Schoenholz, K. B. Aptowicz, M. Schindler, A. C. Maggs, A. J. Liu, and A. G. Yodh. *Phys. Rev. E.*, 88:022315, 2013.

[159] S. S. Schoenholz, A. J. Liu, R. A. Riggleman, and J. Rottler. Understanding plastic deformation in thermal glasses from single-soft-spot dynamics. *Phys. Rev. X*, 4:031014, Jul 2014.

[160] R. L. Jack, A. J. Dunleavy, and C. P. Royall. Information-theoretic measurements of coupling between structure and dynamics in glass-formers. *Phys. Rev. Lett.*, 113:095703, 2014.

[161] J. Gilman. Metallic glasses. *Phys. Today*, 28(5):46–53, 1975.

[162] P. Chaudhari, A. Levi, and P. Steinhardt. *Phys. Rev. Lett.*, 43:1517, 1979.

[163] A. Widmer-Cooper, P. Harrowell, and H. Fynewever. *Phys. Rev. Lett.*, 93:135701, 2004.

[164] A. Widmer-Cooper and P. Harrowell. *J. Chem. Phys.*, 126:154503, 2007.

[165] D. Coslovich and G. Pastore. *J. Chem. Phys.*, 75:784, 2006.

[166] D. Coslovich and G. Pastore. *J. Chem. Phys.*, 127:124504, 2007.

[167] A. Malins, J. Eggers, H. Tanaka, and C. P. Royall. *Faraday Discuss.*, 167:405, 2013.

[168] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, 1995.

[169] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27, 2011.

[170] R. Brüning, D. A. St-Onge, S. Patterson, and W. Kob. *J. Phys.: Condens. Matter*, 21:035117, 2009.

[171] S. Plimpton. *J. Comp. Phys.*, 117:1–19, 1995.

[172] M. L. Falk and J. S. Langer. *Ann. Rev. of Cond. Matt. Phys.*, 2:353, 2011.

[173] V. Mansard, A. Colin, P. Chauduri, and L. Bocquet. *Soft Matter*, 7(5524), 2011.

[174] Jie Lin, Alaa Saade, Edan Lerner, Alberto Rosso, and Matthieu Wyart. On the density of shear transformations in amorphous solids. *Europhys. Lett.*, 105(2):26003, 2014.

[175] Gus LW Hart, Volker Blum, Michael J Walorski, and Alex Zunger. Evolutionary approach for determining first-principles hamiltonians. *Nat. Mater.*, 4(5):391–394, 2005.

[176] Chris J Pickard and RJ Needs. Aluminium at terapascal pressures. *Nat. Mater.*, 9(8):624–627, 2010.

[177] Brian Huskinson, Michael P Marshak, Changwon Suh, Süleyman Er, Michael R Gerhardt, Cooper J Galvin, Xudong Chen, Alán Aspuru-Guzik, Roy G Gordon, and Michael J Aziz. A metal-free organic-inorganic aqueous flow battery. *Nature (London)*, 505(7482):195–198, 2014.

[178] C Austen Angell. Formation of glasses from liquids and biopolymers. *Science*, 267(5206):1924–1935, 1995.

[179] Ludovic Berthier, Giulio Biroli, Jean-Philippe Bouchaud, Luca Cipelletti, and Wim van Saarloos. *Dynamical Heterogeneities in Glasses, Colloids, and Granular Media*. Oxford Scholarships Online, 2011.

[180] Ludovic Berthier and Robert L Jack. Structure and dynamics of glass formers: Predictability at large length scales. *Phys. Rev. E*, 76(4):041509, 2007.

[181] C Patrick Royall, Stephen R Williams, Takehiro Ohtsuka, and Hajime Tanaka. Direct observation of a local structural mechanism for dynamic arrest. *Nat. Mater.*, 7(7):556–561, 2008.

[182] Walter Kob and Hans C Andersen. Scaling behavior in the β-relaxation regime of a supercooled lennard-jones mixture. *Phys. Rev. Lett.*, 73(10):1376, 1994.

[183] Steven Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117(1):1–19, 1995.

[184] Ludovic Berthier and Giulio Biroli. Theoretical perspective on the glass transition and amorphous materials. *Rev. Mod. Phys.*, 83(587):587–645, 2011.

[185] Jerome K. Percus and George J. Yevick. Analysis of classical statistical mechanics by means of collective coordinates. *Phys. Rev.*, 110(1):1–13, 1958.

[186] Anton Smessaert and Jorg Rottler. Structural relaxation in glassy polymers predicted by soft modes: a quantitative analysis. *Soft Matter*, 10:8533–8541, 2014.

[187] A. S. Keys, A. R. Abate, S. C. Glotzer, and D. J. Durian. Measurement of growing dynamical length scales and prediction of the jamming transition in a granular material. *Nat. Phys.*, 3:260–264, 2007.

[188] Paul J Steinhardt, David R Nelson, and Marco Ronchetti. Bond-orientational order in liquids and glasses. *Physical Review B*, 28(2):784, 1983.

[189] Katharina Vollmayr-Lee. Single particle jumps in a binary lennard-jones system below the glass transition. *J. Chem. Phys.*, 121(4781), 2004.

[190] Michael L Falk and James S Langer. Deformation and failure of amorphous solidlike materials. *Annual Reviews of Condensed Matter Physics*, 2:353–373, 2010.

[191] Jun Ding, Sylvain Patinet, Michael L Falk, Yongqiang Cheng, and Evan Ma. Soft spots and their structural signature in a metallic glass. *Proc. Natl. Acad. Sci.*, 111(39):14052–14056, 2014.

[192] Sven Wijtmans and M Lisa Manning. Disentangling defects and sound modes in disordered solids. *arXiv preprint arXiv:1502.00685*, 2015.

[193] Michel Tsamados, Anne Tanguy, Fabien Leonforte, and J L Barrat. On the study of local-stress rearrangements during quasi-static plastic shear of a model glass: Do local-stress components contain enough information? *The European Physical Journal E: Soft Matter and Biological Physics*, 26(3):283–293, 2008.

[194] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch. *Phys. Rev. Lett.*, 97:170201, 2006.

[195] Anton Smessaert and J org Rottler. Distribution of local relaxation events in an aging three-dimensional glass: temporal correlation and dynamical heterogeneity. *Phys. Rev. E*, 88(022314):022314, 2013.

[196] X. Yang, R. Liu, M. Yang, W.-H. Wang, and K. Chen. Structures of local rearrangements in soft colloidal glasses. 2016. arXiv:1601.05919. arXiv.org e-Print archive. http://arxiv.org/abs/1601.05919 (accessed Apr 18, 2016).

[197] J. Helfferich, I. Lyubimov, D. Reid, and J. J. de Pablo. Inherent Structure Energy is a Good Indicator of Molecular Mobility in Glasses. 2016. arXiv:1604.01077. arXiv.org e-Print archive. http://arxiv.org/abs/1604.01077 (accessed Apr 18, 2016).

[198] Sebastian Golde, Thomas Palberg, and Hans Joachim Schope. Correlation between dynamical and structural heterogeneities in colloidal hard-sphere suspensions. *Nat. Phys. [Online early access]*. DOI:10.1038/nphys3709 Published Online: Mar 28, 2016. http://www.nature.com/nphys/journal/vaop/ncurrent/full/nphys3709.html (accessed Apr 18,2016).

[199] Anton Smessaert and Jörg Rottler. Structural relaxation in glassy polymers predicted by soft modes: a quantitative analysis. *Soft matter*, 10(42):8533–8541, 2014.

[200] SP Pan, SD Feng, JW Qiao, WM Wang, and JY Qin. Correlation between local structure and dynamic heterogeneity in a metallic glass-forming liquid. *Journal of Alloys and Compounds*, 664:65–70, 2016.

[201] Jennifer M. Rieser, Carl P. Goodrich, Andrea J. Liu, and Douglas J. Durian. Divergence of voronoi cell anisotropy vector: A threshold-free characterization of local structure in amorphous materials. *Phys. Rev. Lett.*, 116:088001, Feb 2016.

[202] NN Medvedev and Yu I Naberukhin. Shape of the delaunay simplices in dense random packings of hard and soft spheres. *J. Non-Cryst. Solids*, 94(3):402–406, 1987.

[203] Walter Kob, Sandalo Roldan-Vargas, and Ludovic Berthier. Non-monotonic temperature evolution of dynamic correlations in glass-forming liquids. *Nat Phys*, 8(2):164–167, 02 2012.

[204] Walter Kob, Claudio Donati, Steven J Plimpton, Peter H Poole, and Sharon C Glotzer. Dynamical heterogeneities in a supercooled lennard-jones liquid. *Phys. Rev. Lett.*, 79(15):2827, 1997.

[205] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[206] Laurens van der Maaten. Barnes-hut-sne. *arXiv preprint arXiv:1301.3342*, 2013.

[207] Josh Barnes and Piet Hut. A hierarchical O(N log N) force-calculation algorithm. *Nature*, 324(6096):446–449, 1986.

[208] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.